

Sveučilište u Zagrebu, Filozofski fakultet, Ivana Lučića 3

Alenka Kavčič Čolić

**ARHIVIRANJE SLOVENIKE NA INTERNETU
METODOLOŠKI PRISTUP**

Doktorska disertacija

Mentor: prof. dr. Aleksandra Horvat

Zagreb, 2007.

Iskreno se zahvaljujem

mentorici prof. dr. Aleksandri Horvat za korisne i vrlo konstruktivne savjete prilikom pisanja doktorskoga rada,

upravi Narodne i sveučilišne knjižnice u Ljubljani, koje mi je omogućilo znanstveno usavršavanje i izradbu doktorske disertacije,

mojim najdražima, Petru, Marku i Sebastianu, jer su me sve vrijeme istraživanja i pisanja podupirali.

1. UVOD

1.1. Arhiviranje mrežne građe

Terry Kuny, savjetnik u kanadskoj nacionalnoj knjižnici, zapisao je 1998. godine da se upravo nalazimo usred "mračnog srednjeg vijeka digitalnoga doba", jer će većina suvremenoga znanja, kodiranoga i zapisanoga u elektroničkom obliku biti zauvijek izgubljena. Zato je uloga današnjih knjižničara veoma slična ulozi nekadašnjih samostanskih opata, tj. na njima je da sačuvaju što više elektroničkih zapisa.¹

U Kunyjevoj tvrdnji ima puno istine. Djela nastala na materijalnim (analognim) medijima zacijelo će doživjeti 50, 100, a možda i više godina. Kod digitalnih je djela teško predvidjeti koliko će trajati jer je njihov životni ciklus vrlo kratak. Naime, vezana su uz određen stupanj tehnološkoga razvoja i moguće je da za 5 godina neće biti dostupna.

Unatoč kratkotrajnosti elektroničkih publikacija, njihova produkcija i broj objava na mreži iz godine u godinu dramatično se povećava. Istraživači projekta *Web Characterization Project* (<http://www.oclc.org/research/projects/archive/wcp/>) utvrdili su da se broj domena povećao sa 1.313.000 u 1993. godini na 171.638.297 u 2003. godini, što je oko 50% porasta godišnje i više od 13.000% ukupnog porasta u desetih godina. Statistički podatci iz spomenutoga istraživanja pokazuju da se posljednjih godina otvara sve manje novih mrežnih stranica, iako glavni uzrok tome nije manje korištenje ovoga medija, već sve kraći životni ciklus mrežnih stranica. Oko 40% mrežnih publikacija ne doživi godinu dana, a 20% njih u istom se razdoblju uopće ne mijenja. Prosječna dob jedne mrežne stranice je nešto više od mjesec dana.

S informacijskom pismenošću povećava se i broj korisnika interneta. Statističke analize na mrežnim stranicama projekta *Raba Interneta v Sloveniji* (<http://ris.org>) pokazuju da se u ožujku 2005. godine više od 20% slovenske populacije (oko 400.000 osoba) svakodnevno koristilo

¹ "We are, to my mind, living in the midst of digital Dark Ages; consequently, much as monks of times past, it falls to librarians and archivists to hold to the tradition which reveres history and the published heritage of our times." Zapisao: Kuny, T. The digital dark ages? Challenges in the preservation of electronic information // International preservation news. 17, May (1998). URL: <http://www.ifla.org/VI/4/news/17-98.htm#2> (15.12.2006.).

internetom. Od uvođenja interneta u Sloveniji oko 1996. godine bilježi se porast potencijalnih autora na elektroničkom mediju od 5 % godišnje te se stvaranje i udvostručavanje informacija dramatično povećava. Istraživanja na projektu američkoga Sveučilišta Berkeley, *How much information*², pokazuju zabrinjavajuće podatke o količini svih stvorenih informacija u digitalnom obliku na svijetu. Pritom se postavlja pitanje, koje od svih tih informacija vrijedi sačuvati, odnosno koji će od svih trenutačno objavljenih sadržaja na mreži biti glavni izvor informacija našim potomcima o našoj kulturi, povijesti, znanosti ili drugim aspektima, koje su odraz našega suvremenoga civilizacijskoga identiteta.

Većina nacionalnih knjižnica i drugih arhivskih ili depozitarnih ustanova u svijetu svjesna je nužnosti pohrane mrežnih stranica i ostalih izdanja na internetu. Današnji mrežni arhivi, koji su se počeli graditi prije nekoliko godina, postali su bitan izvor različitih istraživanja koja u prošlosti nije bilo moguće predvidjeti. Mrežna građa koja je bila selektivno prikupljena i sačuvana, dostupna je još i danas te omogućava različite komparativne analize na osnovi određene teme ili događaja (naprimjer američki izbori, etnološke zbirke i ostalo). S druge strane, mrežna građa koja je bila automatski prikupljena na osnovi najviše domene ili kakva drugog parametra, nudi različite mogućnosti istraživanja, kao što su informacijsko-tehnološke pozadine mrežnih stranica, kulturne i sociološke analize sadržaja mrežnih stranica i drugo. Dio te građe je u trenutku prikupljanja irelevantan (naprimjer oglasi na internetu ili mrežna trgovina), međutim s vremenom može postati vrlo bitan izvor informacija.³

Istraživanje interneta danas je dio obrazovnih programa na mnogim američkim sveučilištima i mnogi stručnjaci⁴ smatraju to područje novom znanstvenom disciplinom, koja ima svoje znanstvenike, znanstvene publikacije, znanstvene skupove i društva. Međutim, kao nova znanstvena disciplina, vezana uz sasvim novu tehnologiju, još nema dovoljno izrađene teorije i

² How much information 2003. URL: <http://www2.sims.berkeley.edu/research/projects/how-much-info/> (15.12.2006.).

³ Thelwall, M.; Vaughan, L. A fair history of the Web? Examining country balance in the Internet Archive. // Library & Information Science Research. 26(2004), str. 162-176.

⁴ Među njima su: Shrum, W. Internet indiscipline: two approaches to making a field. // The Information Society. 21 (2005), str. 273-275; Monberg, J. Science and technology studies approaches to internet research. // The Information Society. 21 (2005), str. 281-284.

metodologiju i nije organizirana kao znanost. Stoga neki autori⁵ odbijaju spomenute teze i tvrde da je istraživanje interneta alat⁶ za istraživanje ili istraživačka mreža.⁷ U svakom slučaju, istraživanje interneta odnosno internetska istraživanja sve će se više razvijati i mrežni će arhivi pritom odigrati veliku ulogu.

Pri dugoročnoj pohrani mrežne građe razne ustanove koriste se različitim strategijama ili pristupima. Danas je još rano reći koja je strategija bolja, odnosno koji je pristup bolji od ostalih, jer svi imaju svoje prednosti i nedostatke. Dosad su bile najuspješnije one ustanove koje su kombinirale više pristupa. Ipak, uspješnost dugoročne pohrane ovisi ne samo o tehnikama i metodama prikupljanja, nego i o procesu koji im slijedi.

U Bibliotekarskom istraživačkom centru Narodne i sveučilišne knjižnice (u daljem tekstu NUK) već nekoliko godina pokušavamo pronaći rješenje za prikupljanje i dugoročnu pohranu slovenskih mrežnih publikacija. Od 2001. godine pratimo razvoj na tom području, održavamo kontakte s različitim stručnjacima u ostalim nacionalnim knjižnicama i sudjelujemo na različitim međunarodnim znanstvenim skupovima. Izravno smo se upoznali i s praksom Nacionalne knjižnice Australije te već nekoliko godina zaredom sudjelujemo i na radionicama IAWW (*International Web Archiving Workshop*), koje su glavni forum za izmjenu iskustva i informacija iz područja prikupljanja i građe web arhiva. Godine 2004.,⁸ a potom i 2005.,⁹ predstavili smo prva iskustva NUK-a u prikupljanju slovenskoga interneta. Iz radionice je nastala vrlo velika mreža stručnjaka, koji su za NUK bitan izvor informacija i velika pomoć u razvoju mrežnih arhiva.

⁵ Baym, N. K. Introduction: Internet research as it isn't, is, could be, and should be. // *The Information Society*. 21 (2005), str. 229-232;

⁶ Baron, N. S. Who wants to be a discipline? // *The Information Society*. 21 (2005), str. 269-271.

⁷ Hine, C. Internet research and the sociology of cyber-social-scientific knowledge. // *The Information Society*. 21 (2005), str. 239-248.

⁸ Kavčič-Čolić, A. ; Grobelnik, M. Archiving the Slovenian web : recent experiences. // *Proceedings / 4th international web archiving workshop (IAWW04)*, held in conjunction with the 8th European conference on research and advanced technologies for digital libraries, September 16 2004, Bath, UK. Bath : University, 2004. URL: <http://www.iwaw.net/04/index.html>. (15.12.2006.)

⁹ Mladenčić, D. ; Grobelnik, M. ; Kavčič-Čolić, A. Initiatives to preserve Slovenian digital heritage // *Innovation and knowledge economy: issues, applications, case studies / edited by Paul Cunningham and Miriam Cunningham*. Amsterdam [etc.] : IOS Press, 2005. (Information and communication technologies and the knowledge economy). Str. 993-998.

Najviše iskustva na području arhiviranja mrežnih publikacija dobili smo putem različitih projekata koje smo izveli u suradnji s Institutom "Jožef Stefan" (u daljem tekstu IJS), a koje je financirala slovenska Vlada. Rezultati tih projekata bit će opisani u posebnom poglavlju.

1.2. Cilj rada

Od 1995. godine očekivali smo u Sloveniji prihvaćanje novoga zakona o obveznom primjerku, koji bi obuhvaćao i mrežna izdanja. To se napokon dogodilo 2006. godine i NUK je po svojoj funkciji preuzeo najvišu odgovornost za pohranu te građe. U Sloveniji nijedna druga institucija ne pohranjuje dugoročno i sustavno mrežnu i elektroničku građu, što ne znači da ne bi mogla u budućnosti. Stoga su ciljevi doktorske disertacije sljedeći:

- Naći optimalan način za prikupljanje i pohranu elektroničkih publikacija, koje su objavljene na internetu i dio su slovenskoga kulturnoga naslijeđa, jer će te inače biti izgubljena za budućnost.
- Prilagoditi različite međunarodne metodologije i izraditi metodologiju za arhiviranje slovenike na mreži, koja bi omogućila sustavan i organiziran pristup njezinu arhiviranju;
- Predstaviti bitne aspekte arhiviranja slovenskih publikacija na mreži, koje mogu pridonijeti odlučivanju kod odabira optimalnih rješenja.

Disertacija će pokušati dokazati dvije teze:

1. Moguće je izraditi metodologiju i metodološki model koji će usmjeriti knjižničare pri uspostavljanju sustava za arhiviranje mrežne građe.
2. Metodološki model mora biti usklađen s referentnim modelom OAIS.

Pod metodologijom podrazumijevamo skup metoda i postupaka čiji je cilj sistematizacija i organizacija procesa i djelatnosti vezane uz te procese pri stvaranju arhivskoga sustava za mrežnu građu.

U okviru prve teze želimo naći odgovore na sljedeća pitanja:

- Postoji li optimalni pristup za prikupljanje mrežnih publikacija?
- Kakvi su osnovni uvjeti za izradu mrežnog arhiva?
- Je li moguće tretirati mrežne publikacije prema istim kriterijima kao i publikacije na materijalnim nositeljima?
- Je li moguće mrežne publikacije bibliografski opisati jednako kao i publikacije na materijalnim nositeljima?
- Kakve su posljedice genoloških istraživanja kod uvođenja novih tipologija publikacija?
- Kako osigurati njihovo pronalaženje u arhivskom sustavu?
- Kako riješiti sve probleme, koji nastaju zbog autorskog i drugih prava?

U drugoj tezi pokušava se implementirati model korištenjem standarda OAIS (ISO 14721), koji je zacrtao osnovni referentni model pri izradi elektroničkih arhiva. U disertaciji ćemo pokazati je li moguće spomenuti model prenijeti u praksu prikupljanja i arhiviranja mrežne građe. Rezultate tog istraživanja temeljimo na praksi NUK Slovenije. Pritom su prikazani i ocijenjeni rezultati istraživačkih projekata iz toga područja.

1.3. Metodologija

U disertaciji se koristi komparativna analiza različitih rješenja na području arhiviranja mrežnih stranica u Sloveniji i u svijetu. Dragocjena su i iskustva naših susjeda u Nacionalnoj i sveučilišnoj knjižnici u Zagrebu. Također su analizirana rješenja i literaturu iz bitnih područja u vezi prikupljanja, obrade i dugoročnog pohranjivanja mrežne građe. Pri određivanju metodologije i posebice pri definiranju kriterija selekcije i metoda prikupljanja publikacija na mreži, korištena su iskustva na različitim nacionalnim i međunarodnim istraživačkim projektima u NUK-u, kao i informacije dobivene na raznim međunarodnim znanstvenim skupovima. Pri traženju taksonomije i definicije različitih tipologija dokumenata na mreži, korišteni su rezultati međunarodnih istraživanja na području genologije interneta i o karakteristikama mrežnih publikacija. Za izradu metodološkoga modela za arhiviranje slovenike na mrežnim stranicama korišten je funkcionalni model OAIS, koji opisuje entitete i njihove procese u arhivskom sustavu. Metodološki model koji je oblikovan u radu rezultat je prilagođavanja i konkretizacije tih procesa u slovenskoj praksi.

1.4. Organizacija rada

U ovom radu, nakon kraćega uvoda u predmet istraživanja, opisa upotrijebljene metodologije i popisa korištenih termina, prikazani su osnovni koncepti na području arhiviranja mrežnih publikacija. To su digitalni i web arhivi te alati i metode prikupljanja i pohrane mrežne građe. U nastavku je prikazan problem formiranja tipologije i definiranja kriterija odabira. Posebno poglavlje posvećeno je autorskom pravu na području arhiviranja mrežnih publikacija, a u nastavku se obrađuje problematika bibliografskoga opisa. Kad govorimo o arhiviranju, perspektiva dugoročne pohrane od velike je važnosti. O tome govorimo u sljedećem poglavlju, u kome je prikazan referentni model OAIS, metapodatci za dugoročnu pohranu i različite strategije. Dugotrajni identifikatori vrlo su važna informacija za lokalizaciju elektroničke građe i zato im je posvećeno posebno poglavlje. Glavni dio čini nacrt modela procesa u sustavu za arhiviranje mreže, prikaz iskustva u NUK-u i ocjena rezultata projekata.

1.5. Terminologija

Kad je riječ o mrežnoj građi, često zalazimo u interdisciplinarno područje između bibliotekarstva i informacijskih znanosti. Višegodišnja iskustva u suradnji s istraživačima iz područja računalnih znanosti pokazuju da često obrađujemo iste stvari, a da toga nismo svjesni, jer se koristimo sasvim drugom terminologijom i različitim pristupima. Kad smo u NUK-u započeli suradnji s IJS, nakon nekoliko zajedničkih sastanaka ustanovili smo da moramo prije početka projekta zajedno sjesti i usuglasiti zajedničku terminologiju. Primjerice, za njih su naši deskriptori indeksi, digitalni objekti odgovaraju intelektualnim cjelinama, koje nazivamo publikacije, entiteti su različiti pojmovi oko kojih se stvaraju procesi itd. Pošto je i ovdje sličan slučaj, prije početka rada neophodno je objasniti nazive kojima smo se koristili u ovome radu. Mnogi će se bibliotekari teško složiti s nekim terminima jer su uzeti iz tehničkih područja, prije svega iz sustavne analize. Međutim, teško je razumjeti elektroničku mrežnu građu bez tehničkih svojstava i sustavne analize.

Za određivanje koncepta iz područja elektroničke građe na internetu, u ovom se radu služimo terminima iz hrvatskih prijevoda ISBD(ER)¹⁰ ili ISBD(CR).¹¹ Standardi također prihvaćaju izraze "web stranica" za "mrežnu stranicu" i "web mjesto" za "mrežno mjesto". Iako se u hrvatskoj računalnoj literaturi upotrebljava engleska varijanta, odlučili smo se za upotrebu pohrvaćene inačice "mrežna stranica" i "mrežno mjesto".

Gore navedeni standardi definiraju mrežnu stranicu kao stranicu "nekog hipertekstovnog dokumenta na World Wide Web-u. Mrežne stranice zajedno s podskupom naslovnica čine ogromnu zbirku dokumenata koji čine World Wide Web" (ISBD(CR), str. 17), a mrežno mjesto kao: "Odredište World Wide Web-a, identificirano URL-om koje omogućuje pristup i korištenje mrežnih stranica koje se na njemu čuvaju." (ISBD(CR), str. 18).

Za određivanje početne ili naslovne stranice (*home page*) rabi se izraz "naslovnica" iz standarda ISBD(ER) u kome je definirana kao: "Glavni ili početni zaslon hipertekstovnog dokumenta na mrežnome mjestu. Naslovnice su podskup mrežnih stranica i daju obavijesti o sustavima, službama i proizvodima. Pomoću veza u obliku riječi, URL-a i sl., omogućuju pristup drugim dokumentima i mrežnim mjestima." (str. 19)

Iz ISBD(CR) je također preuzet termin "neomeđena građa" (*continuing resources*), u značenju "Jedinice građe koja se izdaje tijekom nekog razdoblja bez unaprijed utvrđenog kraja izlaženja. Neomeđena građa obuhvaća serijske publikacije i tekuću integrirajuću građu." (str. 18), te izraz "integrirajuća građa", koja se u standardu definira kao "Jedinica građe koja se nadopunjuje ili mijenja osuvremenjivanjima koja ne ostaju zasebna, već su uklopljena u cjelinu. Integrirajuća građa može biti omeđena ili neomeđena." (str. 16)

¹⁰ ISBD(ER) : međunarodni standardni bibliografski opis elektroničke građe : prerađeno izdanje ISBD(CF)-a: Međunarodnog standardnog bibliografskog opisa računalnih datoteka / preporučila Radna grupa za pregled ISBD(CF)-a ; [s engleskog prevele, hrvatske primjere odabrale i izradile Tanja Buzina i Sofija Klarin]. Zagreb : Hrvatsko knjižničarsko društvo, 2001. (Povremena izdanja Hrvatskog knjižničarskog društva. Novi niz ; knj. 3)

¹¹ ISBD(CR) : međunarodni standardni bibliografski opis serijskih publikacija i druge neomeđene građe : prerađeno izdanje ISBD(S): Međunarodnog standardnog bibliografskog opisa serijskih publikacija / [priređivač] Međunarodni savez knjižničarskih društava i ustanova ; [s engleskog prevela Ana Barbarić]. Zagreb : Hrvatsko knjižničarsko društvo, 2005. (Povremena izdanja Hrvatskog knjižničarskog društva, ISSN 1334-7667. Novi niz ; knj. 11)

Analogno mrežnim stranicama i mrežnim mjestima upotrebljavamo i termin "mrežna građa", za označavanje svih mogućih varijanti formalne i neformalne građe koja je dostupna na mreži. U istom značenju koristimo i termin "elektroničke publikacije na mreži" ili "građa na mreži", "mrežni dokument", "digitalni objekt", ovisno o perspektivi gledanja. Digitalni je objekt bilo koja zaključena intelektualna cjelina elektroničkog zapisa, bez obzira na to iz koliko elemenata je sastavljen.¹²

Spomenuti standardi prihvaćaju englesku varijantu internetske službe "World Wide Web". U ovom radu prevodimo taj termin kao "svjetsku mrežu" ili samo "mrežu", analogno sa slovenskim prijevodima. Međutim, da bismo spriječili nesporazume, umjesto naziva "mrežni arhiv", u značenju repozitorija za prikupljenu građu na mreži, služimo se engleskom varijantom "web arhiv".

Termin "arhiv" korišten je u značenju depozitarne ustanove ili širega organizacijskoga sustava koji ima zadatak da pohranjuje elektroničku građu za budućnost. Njegovo spremište je repozitorij koji ima ulogu informacijsko-tehnološke aplikacije za pohranu građe.

Za označavanje aplikacije za prikupljanje građe koristimo naziv "robot" ili "aplikacija za prikupljanje", odnosno "sustav za prikupljanje ili preuzimanje mreže", iako postoje različiti asocijativni izrazi kao što su "pauk", "kombajn", "prikupljač" itd.

Za proces prikupljanja mrežne građe upotrebljavamo nazive: "prikupljanje", "preuzimanje", "skupljanje".

Za proces otvaranja i čitanja elektroničke građe rabimo nazive "dostup" ili "pristup".

Za označavanje procesa i aktivnosti koje su u engleskom jeziku definirane kao *long-term preservation*, koristimo termin "dugoročna pohrana". Analogno tomu, metapodatke koji nose potrebne informacije za taj proces nazivamo "metapodatci za dugoročnu pohranu".

¹² Stančić za digitalni objekt koristi naziv elektronički informacijski objekt i definira ga kao "onaj objekt koji je nastao uz pomoć informacijske tehnologije, bez obzira je li to njegov izvorni oblik ili je riječ o gradivu u klasičnom obliku prenesenom u elektroničku okolinu." (str. 28). Vidi djelo: Stančić H. Očuvanje elektroničkih informacijskih objekata : arhivi, knjižnice, muzeji - zajednička koncepcija. // 7. Seminar arhivi, knjižnice, muzeji..., 2004. Str. 26-35.

2. KARAKTERISTIKE PUBLIKACIJA NA MREŽI

Internet je najveći svjetski komunikacijski i publikacijski kanal.¹³ Nepogrešiv je u informacijskom društvu i vrlo se brzo širi u razvijenom svijetu, kao i u državama u razvoju. Javnost često miješa naziv internet sa svjetskom mrežom (*www*), ne znajući da je svjetska mreža samo dio interneta. Prema definiciji ISBD(ER), internet je "Globalna računalna mreža koja se sastoji od mnogo manjih mreža međusobno povezanih internetskim protokolom (IP) i drugim sličnim protokolima" (str. 17). Svjetska mreža jedna je od mnogih internetskih službi "koja povezuje dokumente pomoću hiperteksta. Veze u obliku riječi, URL-a itd. omogućuju pronalaženje i pristup dokumentima na internetu" (ISBD(ER), str. 24). U usporedbi s donedavno prevladavajućim tiskanim oblicima, elektronički dokumenti imaju sasvim druga svojstva. Njihovi sadržaji nisu dostupni bez informacijske tehnologije, koja se neprestano razvija i uzrokuje pojavu velike raznolikosti računalnih formata koji relativno brzo zastarijevaju. Njihovi materijalni nositelji također su vrlo nestabilni i njihov životni ciklus mnogo je kraći od izdanja tiskanih na papiru. Osim toga, zahtijevaju posebne uvjete za skladištenje, kao što je prilagođeni sustav za pohranu i upravljanje elektroničke građe – repozitorij, i specifičnu praksu obrade i upravljanja elektroničkim publikacijama.

Većina elektroničkih publikacija, koje obrađujemo u knjižnicama, objavljena je na svjetskoj mreži. Međutim, znamo li doista kakve su te publikacije? Možemo li ih tretirati jednako kao i tradicionalnu građu na materijalnim nositeljima? Dokumenti na mreži donose nam novu dimenziju koju u tradicionalnoj građi ne nalazimo. Dok kod knjiga, kao i druge građe tiskane na papiru, pratimo sadržaj linearno (riječi čitamo od početka do kraja bez prekida), u digitalnom mediju susrećemo se s novom trodimenzionalnom logikom. Koncepti kao što su "mrežna lokacija stranice građe", "hijerarhijska struktura njezinih sastavnih dijelova", "razgranat sustav poveznica na mreži", "cjelovitost" i "dostupnost do sadržaja", otvaraju nam nov način shvaćanja pisanih sadržaja u digitalnom okruženju. Upravo ta nova dimenzija otvara polemiku o novim žanrovima ili tipologiji mrežne građe.

¹³ Po Masanèsu "Internet je izgrađen na nekoliko osnovnih protokola (primjerice TCP/IP) i sustava (primjerice DNS), koji omogućavaju jednostavno povezivanje računala u mrežu." Publiciranje na mreži omogućuju standard URI (*Universal Resource Identification*), HTTP (*Hypertext Transport Protocol*) za preuzimanje i dostavu dokumenata i HTML (*Hypertext Mark-up Language*) (str. 79-80). Vidi: Masanès, J. Web archiving // Digital preservation. / uredili Deegan, M. ; Tanner, S. London : Facet Publishing, 2006. Str. 78-97.

S pojavom publikacija u elektroničkom obliku i sa selidbom publikacija na digitalno okruženje u informacijskom društvu sasvim se promijenilo značenje publikacije, njezine dostupnosti, reprodukcije i distribucije.^{14,15} Digitalna priroda elektroničke publikacije omogućava njezino jednostavno i brzo umnožavanje, a pritom su kopije identične originalu. Njihova reprodukcija također je veoma pojednostavljena jer tisuću ljudi može istovremeno pristupiti samo jednom određenom primjerku. Za razliku od tiskane građe, elektronička publikacija preskače sve granice tradicionalnoga nakladništva, te se uvode nove pravne uredbe, kojima se pokušava kontrolirati nedopušten pristup i umnožavanje.

Koncept publikacije i njezina publiciranja također se mijenjaju. Danas svatko može objaviti elektroničku publikaciju bez recenzije, mijenjati je bez ikakvih obavijesti ili je skinuti s poslužitelja. Usto, vrijeme trajanja te mrežne publikacije svaki je put sve više ograničeno.

Iz informacijske perspektive internet omogućava vrlo brzu diseminaciju informacija, znanstvenih dostignuća i istraživanja. Međutim, nisu točna predviđanja da će publikacije u digitalnom obliku potisnuti tradicionalno nakladništvo. Danas se događa da se većina građe udvostručava, mnoge publikacije koje su bile objavljene na papiru digitaliziraju se i objavljuju na internetu. Relevantne informacije, izvještaji i različiti oblici sadržaja, kao što su doprinosi na znanstvenim skupovima, umnožavaju se i objavljuju ne samo na različitim poslužiteljima, već i na različitim formatima (CD-ROM, internet, papir...).

Znači, nalazimo se u situaciji zasićenosti informacijama. Internet u većoj mjeri doprinosi njihovu udvostručavanju. Sve se više objavljuje u elektroničkom obliku. Unatoč veoma velikim

¹⁴ The Digital Dilemma: Intellectual Property in the Information Age. // Committee on Intellectual Property Rights and the Emerging Information Infrastructure, the Computer Science & Telecommunications Board, and the Commission on Physical Sciences, Mathematics, and Applications. Washington : National Academy Press, 2000. URL: http://bob.nap.edu/html/digital_dilemma/. (15.12.2006.) Izvještaj je godine 1997. naručio američki Odbor za Računalništvo i Telekomunikacije (*Computer Science and Telecommunications Board* - CSTB) Nacionalnih akademija SAD-a (*National Academies*) s ciljem da prouči utjecaj novonastale digitalne informacijske infrastrukture na intelektualno pravo.

¹⁵ Samuelson, P. and Davis, R. The digital dilemma: a perspective on intellectual property in the information age, paper written for presentation at the 28th Annual Telecommunications Policy Research Conference, 2000. URL: <http://www.sims.berkeley.edu/~pam/papers/digdilsyn.pdf> (15.12.2006.), p.7.

kapacitetima i sposobnosti sustava za pretraživanje po svjetskoj mreži, treba nam mnogo vremena da izlučimo relevantne informacije koje su nam potrebne. Gulli in Signorini¹⁶ su krajem siječnja 2005. godine ocijenili da suvremeni sustavi za pretraživanje interneta indeksiraju oko 11,5 milijardi javno dostupnih mrežnih stranica. Naravno, različiti pretraživači daju različite rezultate. Naprimjer, u kolovozu 2006. godine Yahoo je našao 2.390.000 slovenskih mrežnih mjesta, Alta Vista 7.680.000, a Google 25.700.000. Dobiveni podatci ukazuju na činjenicu da je gotovo nemoguće točno izračunati količinu svih mrežnih stranica ili mrežnih mjesta u svjetskoj mreži.

Zbog složenosti problematike, mjerenje količine i utvrđivanje karakteristika mrežnih stranica predmet su istraživanja različitih projekata. U Sloveniji je takav projekt RIS (*Raba Interneta v Sloveniji*), čiji je glavni cilj istraživanje korištenja internetom među slovenskim stanovništvom. Na njihovim stranicama (www.ris.org) mogu se naći statistički podatci iz različitih izvora, između ostalog saznajemo da je u siječnju 2006. godine u svijetu bilo 394.991.609 "hostova" (Vir: *Internet Systems Consortium*, <http://www.isc.org>), a u ožujku te iste godine 77.568.868 mrežnih mjesta (Vir: *Netcraft*, <http://news.netcraft.com>). Po podacima ARNES-a¹⁷ na spomenutim stranicama RIS saznajemo da je u ožujku 2006. godine u Sloveniji bilo 39.197 registriranih domena na *.si. Broj registriranih domena znatno raste od 1996. godine. Granica od 10.000 registriranih domena prijeđena je već 2001. godine, a nakon tri godine taj se broj udvostručio. Najveći porast zabilježen je 2005. godine, kada je registracija omogućena svim pravnim osobama - prije je domena *.si bila dodijeljena samo neprofitnom sektoru, posebice akademskim i drugim javnim ustanovama i istraživačima. U Sloveniji je u studenom 2005. godine već bilo 13 milijuna dokumenata na mreži (Vir: *Najdi.si*, www.najdi.si), dok je u siječnju 2006. godine bilo registrirano 59.790 "hostova" (Vir: *RIPE NCC*, <http://www.ripe.net/>).

Među većim međunarodnim projektima koji su pratili rast i razvoj interneta u svijetu je *Web Characterization Project* (www.oclc.org/research/projects/archive/wcp/stats/size.htm) konzorcija OCLC. Na njihovim stranicama možemo pronaći statističke podatke o rastu broja mrežnih mjesta od 1998. do 2003. godine. Mrežno mjesto definirano je kao "razgovijetna lokacija na Internetu,

¹⁶ Gulli, A. ; Signorini, A. The indexable web is more than 11.5 billion pages. // ACM URL: <http://www.cs.uiowa.edu/~asignori/pubs/web-size/> (15.12.2006.)

¹⁷ *Academic Research Network of Slovenia* - slovenska akademska istraživačka mreža, institucija koja je među prvima u Sloveniji započela s registracijom domena. Dostupna na URL: www.arnes.si (15.12.2006.)

koju identificira njena IP adresa i koja na zahtjev protokola HTTP šalje odgovor sa kodom 200 i prikazuje odgovarajuću mrežnu stranicu. Sastavljeno je iz svih povezanih mrežnih stranica na istoj IP adresi."¹⁸ U **tabeli br. 1** vidi se da je stupanj rasta mrežnih mjesta sve manji. Glavni razlog tomu ne treba tražiti u sve manjoj produkciji na mreži, već u sve većoj promjenljivosti mrežnih mjesta, brzom razvoju informacijske tehnologije, što utječe na zamjenu poslužitelja, odnosno prestanak njihova djelovanja, što smanjuje životni ciklus mrežnih mjesta.¹⁹ (<http://www.oclc.org/research/projects/archive/wcp/stats/misc.htm>). U **tabeli br. 2** prikazani su postotci preživjelih mrežnih mjesta po godinama.

Tabela br. 1: Statistički podatci o broju mrežnih mjesta 1998.-2002. (Vir: WCP, OCLC)

Podatci iz godine:	Broj mrežnih mjesta	Stupanj rasta br. mrežnih mjesta
1998.:	2.851.000	
1999.:	4.882.000	71%
2000.:	7.399.000	52%
2001.:	8.745.000	18%
2002.:	9.040.000	3%
1998.-2002.:		217%

Tabela br. 2: Postotak preživjelih mrežnih mjesta po godinama

	1998.	1999.	2000.	2001.	2002.
1998.	100%	56%	35%	25%	13%
1999.	NP	100%	55%	37%	19%
2000.	NP	NP	100%	55%	33%
2001.	NP	NP	NP	100%	51%
2002.	NP	NP	NP	NP	100%

¹⁸ A Web site is defined as a distinct location on the Internet, identified by an IP address, that returns a response code of 200 and a Web page in response to an HTTP request for the root page. The Web site consists of all interlinked Web pages residing at the IP address." URL: www.oclc.org/research/projects/archive/wcp/stats/size.htm (15.12.2006.)

¹⁹ Nestabilnost mrežne građe dokazuju Klarin, Pigac i Pavletić u svom istraživanju 152 naslova hrvatske elektroničke serijske građe na mreži koje su vodili od siječnja 2000. do veljače 2001. Utvrdili su da su u roku od 15 mjeseci dva naslova prestala izlaziti, kod 16 (10.45%) naslova poveznice više nisu djelovale, a 12 (7,84) naslova je promijenilo adresu URL. Jedna od serijskih publikacija promijenila je stvarni naslov. Vidi: Klarin, S ; Pigac, S ; Pavelić, D. Croatian reomte access electronic serials : results of a survey. // International Cataloguing & Bibliographic Control. 30, 4 (2001), str. 70.

Nažalost, projekt je završio 2003. godine i u vrijeme pisanja ovoga rada nije bilo novijih podataka. Podatci projekta iz 2003. godine pokazuju donekle ustaljen trend jednogodišnjega životnoga ciklusa mrežnih mjesta. Iz drugih izvora²⁰ saznajemo da je prosječni životni ciklus jedne mrežne stranice 44 dana. Novija istraživanja²¹ dokazala su da godinu dana preživi 36% mrežnih stranica, a tri godine samo 12%. 70% svih mrežnih mjesta doživi u prosjeku godinu dana. Što se tiče sadržaja, samo 13% ne mijenja se u cijeloj godini. To znači da veći dio naše pismene duhovne, društvene i kulturne baštine ne doživi niti jednu godinu.

Kod tako velike produkcije nema dovoljno vremena za vrednovanje svega što je napisano, a još manje za odlučivanje o tome što bi moglo biti zanimljivo budućim generacijama. Bilo bi potrebno sačuvati što više naše sadašnjosti da bismo u budućnosti mogli imati uvid u svoju prošlost.

Međutim, u želji da sačuvamo mrežnu građu može nastati problem, a to je nemogućnost da neke njezine dijelove prikupimo postojećim alatima. To je problem dubinske ili nevidljive mreže, o kojoj je riječ u sljedećem poglavlju.

²⁰ Collections policy statement : web site capture & archiving. Washington: Library of Congress, 2004. URL: <http://www.loc.gov/acq/devpol/webarchive.html>. (15.12.2006.); Christensen-Dalsgaard, B. ... [et al.]. Experiences and conclusions from a pilot study: web archiving of the district and country elections 2001. Final report for the pilot project "netarkivet.dk", February 2003. <http://netarchive.dk/publikationer/webark-final-rapport-2003.pdf> (15.12.2006.); Lyman, P. ... [et al.]: How much information. URL: <http://www.sims.berkeley.edu/how-much-info/> (15.12.2006.); Mannerheim, J. The WWW and our digital heritage - the new preservation tasks of the library community // 66th IFLA Council and General Conference, Jerusalem, Israel, 13-18 August 2000. URL: <http://www.ifla.org/IV/ifla66/papers/158-157e.htm>. (15.12.2006.)

²¹ Gomes, D. ; Silva, M. On URL and content persistence. Prosinac 2005, str. 5. URL: <http://www.di.fc.ul.pt/tech-reports/05-21.pdf> (15.12.2006.)

2.1. Dubinska ili nevidljiva mreža

Svjetska mreža sastavljena je iz dvaju slojeva, površinske i dubinske mreže.²² Površinska ili vidljiva mreža sadrži mrežne stranice, odnosno mrežna mjesta koja konvencionalni mrežni pretraživači mogu indeksirati. Dubinska ili nevidljiva mreža sadrži informacije koje aplikacije za prikupljanje mrežne građe (*crawleri*), kojima se koriste internetski pretraživači, ne indeksiraju ili nisu normalno dostupne.²³

Bergman²⁴ tvrdi da u početku razvoja interneta i poslije s pojavom tekstova u HTML dubinske mreže ustvari nije ni bilo. Tri čimbenika su 1996. godine utjecala na njezin nastanak: (1) prodor tehnologije baza podataka na internet preko Bluestonovega (danas HP) Sappire/Web i potom Oracla; (2) komercijalizacija mreže s direktorijima i pretraživačima te pojava e-tržišta; (3) web poslužitelji koji su bili prilagođeni "dinamičnim" mrežnim stranicama (naprimjer Microsoftova AS Pand ili Unix PHP tehnologije).

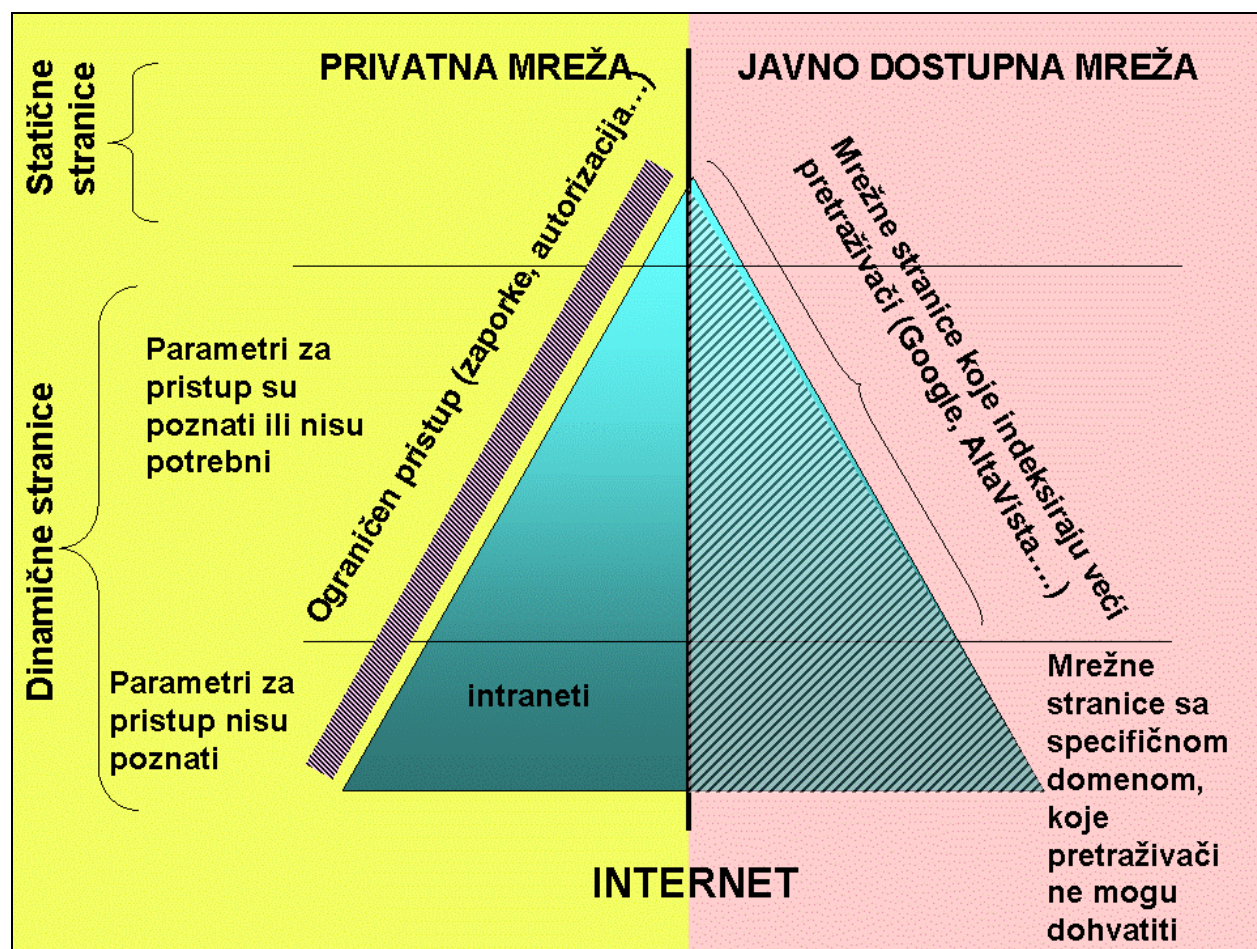
Bergman je na osnovi istraživanja u ožujku 2000. godine za Bright Planet objavio Bijelu knjigu (*White paper*) u kojoj procjenjuje da je dubinska mreža oko 400-550 puta veća od površinske mreže. Pomoću spomenutoga istraživanja prebrojano je 7500 TB informacija u dubinskoj mreži i samo 19 TB u površinskoj mreži. U dubinskoj mreži bilo je 550 milijardi samostalnih dokumenata, a u površinskoj tek jedna milijarda, odnosno oko 200.000 mrežnih mjesta. (vidi **sliku br. 1**). Ovim Bergman zaključuje da je kvaliteta dubinske mreže 1.000-2.000 puta veća od kvalitete površinske mreže. Dubinska mreža ima veći promet od površinske i više od 50% mrežnih mjesta u dubinskoj mreži odnosi se na specifične teme. Zanimljivo je otkriće da je 2000. godine 95% dubinske mreže bilo javno i besplatno dostupno.

²² Laura Turner u svome članku Digging Deeper: The Deep Web (2001) tvrdi da prvi put je dr. Jill Ellsworth godine 1994 upotrijebila frazu "nevidljiva mreža" da bi definirala informacijske sadržaje koji su nevidljivi ili nedostupni konvencionalnim mrežnim pretraživačem. Riječ je pogrešna jer ako ih pretraživači nisu u stanju vidjeti, to ne znači da te informacije nisu dostupne. Dostupne su onima kojima trebaju. Primjerniji izraz je "*deep web*", što bi u prijevodu značilo "dubinska mreža", jer su te informacije duboko skrivene na svjetskoj mreži.

²³ Pedley, P. The invisible web : searching the hidden parts of the internet. London : ASLIB-IMI, 2001.

²⁴ Bergman, M. K. The 'Deep' Web : Surfacing Hidden Value. White Paper. BrightPlanet. Sept. 2001. URL: <http://www.press.umich.edu/jep/07-01/bergman.html> (15.12.2006.)

Pedley²⁵ naziva taj skriveni sloj mreže "jedva vidljiva mreža " i "tamna/neprozirna mreža " (*opaque*). Ona sadrži liječničke baze podataka, diskusijske popise, baze podataka patenata, telefonske brojeve, elektroničku poštu, adrese, vladine baze podataka, znanstvene baze podataka, rječnike, tezauruse, kataloge proizvoda itd.



Slika br. 1: Dubinska mreža. (Po: Baeza-Yates, R. ; Castillo, C. *Crawling the infinite web*)

Razlozi za njezino postojanje su sljedeći:²⁶

- dubina prikupljanja (*crawl*) - pretraživači mogu imati ograničen broj stranica koje mogu indeksirati za svako mrežno mjesto; Bergman²⁷ tvrdi da konvencionalni pretraživači koji imaju najveći broj indeksiranih stranica (npr. Google, Northern light) pretražuju tek 0,03% od svih stranica koje su im dostupne;

²⁵ Pedley, P. op.cit.

²⁶ Pedley, P. op.cit.

²⁷ Bergman, M.K. The 'Deep' Web ... 2001.

- čestota obnavljanja - dok se neka mrežna mjesta svakodnevno obnavljaju, pretraživači ih možda obidu tek svakih nekoliko tjedana ili mjeseci te tako nastaje raspon između novih podataka i indeksiranih informacija;
- moguće je namjestiti pretraživače tako da neka mrežna mjesta uopće ne indeksiraju;
- stranice su promijenile adresu ili poslužitelja;
- vlasnik mrežnoga mjesta nije više zainteresiran za održavanje stranice;
- poduzeće koje je vlasnik mrežne stranice prestalo je poslovati;
- stranica bila namjerno skinuta sa mreže.

Pedley²⁸ navodi vrste mrežnih mjesta koje su skrivene u dubinskoj mreži:

1. baze podataka koje zahtijevaju lozinku za pristup ili ispunjavanje registracijskih obrazaca (besplatne ili uz plaćanje).
2. periodični arhivi;
3. dinamično generirani podatci ili dinamične baze podataka - podatke dobivamo samo pretraživanjem jer su te baze u većini slučajeva sastavljene iz CGI-a (*Common Gateway Interface*) skripta, JavaScripta ili ASP-a; za njih je karakteristično da se u URL-u pojavi znak pitanja što pretraživači odmah razumiju kao prijetnju ili simbol za kraćenje (*truncation*) i zato ih ne možemo indeksirati; tu ulaze kalkulatori, kalendari, različite personalizacije sučelja i slično;
4. CGI skripti su programi napisani u Perlu ili C++ i povezani su s poslužiteljem; pretraživač im ne pristupa izravno, već preko poslužitelja koji prosljeđuje zahtjev CGI-u, a zatim i odgovor pretraživaču;
5. katalozi;
6. *MacroMedia Flash* - Flash je multimedijски kreator i *plug in*;
7. *Streaming media* - sadrži audio, video i druge medijske zapise; obično su nedostupni pretraživačima;
8. *Real time data* – burzovna izvješća, vremenske informacije, vozni redovi vlakova, autobusa, aviona i slično, to su opsežne baze podataka koje se većinom stvaraju dinamično i na zahtjev;
9. baze znanja - zbog velike opsežnosti;

²⁸ Pedley, P. op.cit, str. 25-31.

10. shop-bots - potrošačka uspoređivanja sa popisima cijena, trgovina itd.

Cristensen-Dalsgaard²⁹ nabraja mrežna mjesta koja su uzrokovala teškoće kod prikupljanja danske mrežne građe sa Heritrixom i koje su bile dio dubinske mreže:

- mrežna mjesta zaštićena sa lozinkama;
- obrasci za pretraživanje;
- sustavi koji se osnivaju na različitim transakcijama;
- sustavi koji se osnivaju na interakcijama;
- personalizirana mrežna mjesta;
- beskonačna mrežna mjesta.

Te stranice ne samo da uzrokuju teškoće kod prikupljanja, već to čine i kod dugoročne pohrane, jer je većini digitalnih objekata na zahvaćenim stranicama potreban interpreter.

Dubinska mreža sadrži veoma bitan dio pisane kulturne baštine i zato je potrebno naći način za njezinu pohranu. Koliko ćemo te građe uspjeti prikupiti i pohraniti, ovisi o alatu i metodama za prikupljanje mrežne građe. O tome će biti govora u posebnom poglavlju.

²⁹ Christensen-Dalsgaard, B. Web archive activities in Denmark. // RLG DigiNews. 8, 3 (2004). URL: http://www.rlg.org/en/page.php?Page_ID=17661#article0 (15.12.2006.)

3. DIGITALNI I WEB ARHIVI

3.1. Definicija koncepta

Nastanak elektroničke građe bio je glavni povod za izgradnju digitalnih knjižnica. Oba koncepta nastala su nedavno i još se uvijek razvijaju, u stručnoj terminologiji možemo naći termine koji naglašavaju različite aspekte njihova postojanja i faze njihova razvoja, kao što su "elektronička", "digitalna" ili "virtualna" knjižnica. Automatizacijom knjižnih kataloga nastala je elektronička knjižnica. Ona je sve više uključivala i elektroničku građu. S digitalizacijom su nastale i digitalne zbirke, a time je i knjižnica postala "digitalna". Termin "digitalna knjižnica" danas označava digitalno dostupnu građu i sve se više miješa s konceptom digitalizacije te se upotrebljava i za označavanje zbirke mrežno ili lokalno dostupne digitalizirane građe. "Knjižnica bez zidova" ili "virtualna knjižnica" označava virtualan pristup sadržaju njezinih fondova, što znači da materijalna svojstva nisu bitna. Digitalna knjižnica također sadrži elektroničku građu do koje je uređen virtualni pristup. Sva tri spomenuta termina često se poklapaju ili miješaju. Primjerice, knjižničari Digitalne knjižnice Sveučilišta u Kaliforniji s ponosom tvrde da je njihova knjižnica prva koja materijalno ne postoji, to jest virtualna je.

Za razliku od tradicionalnih knjižnica, u kojima postoji dugogodišnja praksa pri rukovanju i organizaciji tiskane građe, u digitalnoj knjižnici knjižničari se susreću s elektroničkom građom. Ta je građa vezana uz informacijsku tehnologiju i zato zahtijeva posebne uvjete za pohranu. Međutim, i u ovom se slučaju pojavljuju nejasnoće kod naziva skladišta za pohranu elektroničke građe. Prije svega treba razlikovati kakvoj građi i za kakvu pohranu je spremište namijenjeno. Zatim je potrebno sagledati i procese koji se odvijaju u njemu. Obično se rabi termin repozitorij za označavanje bilo kakva skladišta, u kojem se pohranjuju datoteke digitalne građe. Repozitoriji ne podrazumijevaju dugoročnu pohranu te građe i većinom su obične baze podataka koje omogućavaju pronalaženje smještenih datoteka.

Pojam dugoročne pohrane elektroničke građe vezan je uz pojam digitalnoga arhiva. Iako se i tu pojavljuje različito nazivlje: naprimjer "depozitarni sustav za pohranu elektroničkih

publikacija",^{30,31} "elektronički ili digitalni arhiv" te "elektronički ili digitalni repozitorij". U literaturi se često miješaju pojmovi "digitalni arhiv" i "digitalni repozitorij".³² Standard OAIS, o kome će poslije biti govora, rabi naziv "arhiv" za označavanje cjelovitoga sustava za dugoročnu pohranu također elektroničke građe. Zbrka nastaje kada RLG i OCLC objavljuju svoj izvještaj *Trusted Digital Repositories: attributes and responsibilities*,³³ u kome digitalnim repozitorijima nazivaju zapravo isti koncept koji je u spomenutom standardu upotrijebljen za digitalni arhiv. Spomenuti je izvještaj nastao kada se o osnivanju referentnoga modela OAIS već vodila javna rasprava. Pritom treba razumjeti da se u Sjedinjenim Američkim Državama (u daljem tekstu SAD) u to vrijeme govorilo o institucijskim repozitorijima ili skladištima, koji su za druge organizacije, pa i knjižnice, čuvale elektroničku građu.³⁴ Tom su terminu dodali i pojam "pouzdanosti", čime bile izdvojene organizacije koje ozbiljno preuzimaju ulogu skladištenja i dugoročne pohrane elektroničke građe. Pošto su se oba termina u isto vrijeme upotrebljavala za isti koncept, javilo se mnogo nejasnoća i nesporazuma u stručnoj javnosti o tome što zapravo označava repozitorij, a što arhiv. Pod pojmom arhiv aludiralo se na arhivsku instituciju, a pod repozitorijem na druge organizacije. Međutim, u računalnoj terminologiji, repozitorij je spremište. Unatoč različitim upotrebama termina "digitalni arhiv" i "digitalni repozitorij", primjerna je upotreba pojma "arhiv" prema standardu OAIS, u značenju organizacije u kojoj se odvijaju svi potrebni procesi za dugoročnu pohranu elektroničke građe, dok se "repozitorij" u užem značenju odnosi na spremište u smislu aplikacije ili baze podataka, bez kojeg procesi arhiviranja ne bi bili mogući.³⁵

³⁰ Jansen, H. Permanent access : the e-Depot at the Koninklijke Bibliotheek. URL: <http://www.nla.gov.au/webarchiving/JansenHans.rtf> (15.12.2006.)

³¹ Verhoeven, I. H. Archiving web publications. Amsterdam : IBM ; The Hague : Koninklijke Bibliotheek, Dec. 2002. Str. 45. (IBM/KB Long-Term Preservation Study Report Series ; 6) URL: http://www.kb.nl/hrd/dd/dd_onderzoek/reports/6-webpublications.pdf (15.12.2006.)

³² U novijem rječniku za knjižničarstvo i informacijske znanosti nalazimo sljedeću definiciju digitalnoga arhiva: "A system designed for locating, storing, and providing access to digital materials over the long term. A digital archive may use a variety of preservation methods to ensure that materials remain usable as technology changes, including emulation and migration." Dovoljan dokaz o miješanju i izjednačavanju naziva arhiv s repozitorijem dokazuje činjenica da se u istom mjestu navodi digitalni repozitorij kao sinonim digitalnoga arhiva. Vidi: Reitz, J. M. Dictionary for library and information science. Westport : Libraries Unlimited, 2004. Str. 216.

³³ Trusted Digital Repositories: Attributes and responsibilities : an RLG-OCLC Report. Mountain View, CA : RLG, 2002. URL: <http://www.rlg.org/longterm/repositories.pdf> (21.9.2004.)

³⁴ U SAD-u mnogi veći nakladnici elektroničkih serijskih publikacija obavljaju funkciju institucionalnih repozitorija, koji pohranjuju dugoročno ili kratkoročno produciranu građu.

³⁵ Britanski Digital Curation Centre (DCC), koji je vodeća institucija na području dugoročne pohrane elektroničke građe, izjednačava pojmove "institutional repositories", "digital repositories" i "digital archives". URL: <http://www.dcc.ac.uk/resource/technology-watch/dspace>. (15.12.2006.)

Podvrsta digitalnih arhiva su web arhivi, koji obuhvaćaju samo mrežnu građu. Mrežno dostupna građa, kao što ćemo poslije vidjeti, ima drugačiju strukturu i prikuplja se pomoću specijaliziranih alata ili aplikacija (robota prikupljača) koji na poseban način smještaju i pohranjuju tu građu. Web arhivi razlikuju se od ostalih arhiva po načinu organizacije građe, smještaja, pohrane i pristupa elektroničkoj građi.

3.2. Uvjeti za izgradnju pouzdanih digitalnih arhiva

Početkom ovoga stoljeća većina knjižnica i drugih institucija pohranjivala je prilično veliku količinu elektroničke građe. Međutim, još i danas se malo koja od tih institucija brine za dugoročno održavanje pristupa toj građi. Pitanje koje se u to vrijeme pojavilo jest kako zapravo osigurati da sva pohranjena elektronička građa u tim institucijama bude pohranjena za buduće generacije. Koje uvjete te institucije moraju ispunjavati da bismo im vjerovali i povjerali dugoročnu pohranu elektroničke građe, koja je znatan dio kulturne baštine?

RLG i OCLC su u izvještaju *Trusted digital repositories: attributes and responsibilities* iz 2002. godine pokušali odgovoriti na sva ta pitanja. U tom se izvještaju navodi da je glavni zadatak pouzdanoga digitalnoga arhiva (*trusted digital repository*) omogućiti određenoj zajednici u sadašnjosti i budućnosti pouzdan i dugoročan pristup pohranjenoj digitalnoj građi. U istom izvještaju navode se njegove glavne karakteristike, između ostaloga organizacijska, financijska, poslovna i proceduralna pouzdanost u vezi s dugoročnom pohranom građe i usklađenost njegova repozitorija s referentnim modelom OAIS (*Open Archival Information System*).³⁶

Organizacijska pouzdanost znači da organizacija mora biti strateški usmjerena k prikupljanju, arhiviranju i dugoročnoj pohrani elektroničke građe. To znači da organizacija ima kadrove i stručnjake osposobljene za obavljanje funkcije pouzdanoga arhiva, ima dokumentirane smjernice, kao i sve procese i postupke koji su potrebni za njezino djelovanje na tom području. Procesi koji

³⁶ Trusted Digital Repositories: Attributes and responsibilities : an RLG-OCLC Report. Mountain View, CA : RLG, 2002. Str. 13 i 33.

se odvijaju u okviru postavljenih strateških ciljeva ne smiju biti podređeni komercijalnom interesu.

Organizacija koja obavlja funkciju pouzdanoga digitalnoga arhiva mora biti financijski pouzdana. To znači da ima osigurana redovita financijska sredstva za razvoj i održavanje arhiva. Zato su javne ustanove, koje financiraju vlade, idealne za dugoročnu zaštitu kulturne baštine, jer u državnom proračunu moraju biti predviđena sredstva za tu namjenu.

Pouzdan digitalni arhiv odgovoran je za svoje poslovanje, koje je u skladu s dobrom praksom. Mora osigurati pouzdanost i trajnost pomoću primjene domaćih i/ili međunarodnih standarda u vezi s procesima, metodama i infrastrukturom, koja je potrebna za djelovanje repozitorija.

Organizacija koja preuzima ulogu pouzdanoga digitalnoga arhiva mora biti pouzdana i u smislu da jamči visoku sustavnu zaštitu pohranjene elektroničke građe. To znači da mora predvidjeti i planirati različite scenarije u slučaju ispada sustava, gubljenja ili promjene podataka, prirodnih i drugih nesreća te planirati pohranu sigurnosnih kopija građe na lokaciji udaljenoj više od 100 km od lokacije glavnoga arhiva. Mora imati plan za povratak podataka, kao i formalni sukcesijski plan.

Pouzdana digitalni arhivi moraju preuzeti odgovornost za svoje djelovanje, za odabir građe, za upravljanje cijelim životnim ciklusom digitalne građe, za ispunjavanje očekivanja zainteresiranih zajednica, za sređivanje autorsko-pravnih aspekata i za sve izdatke kod dugoročne pohrane. Moraju osigurati zaštitu od upada u sustav, zaštitu autorskih prava i zaštitu od nesreća te se moraju brinuti za dugoročnu pohranu.

Moraju osigurati tehnološku, kadrovsku i proceduralnu primjernost kod dugoročne pohrane elektroničke građe. Digitalni arhiv mora predvidjeti različite strategije za dugoročnu pohranu i u tu svrhu mora osigurati odgovarajuću informacijsko-tehnološku infrastrukturu te detaljan plan postupaka i scenarija za izvođenje navedenih strategija.

Da bi se na neki način formalizirala ta "pouzdanost", pojavili su se i različiti sustavi dodjeljivanja certifikata. RLG i OCLC zauzimaju se za uvođenje sustava certificiranja koji bi se odvijao u više faza i nakon nekoliko godina bi se obnavljao.³⁷ U Njemačkoj je *Deutsche Initiative für Netzwerkinformation* 2003. godine počeo s dodjeljivanjem DINI-eva certifikata³⁸ organizacijama koje su zadovoljile osnovne kriterije za dugoročnu pohranu digitalne građe. Osnova za analizu su upitnici koje zainteresirane arhivske organizacije ispunjavaju i šalju agenciji DINI uz plaćanje troškova.³⁹

3.3 Nacionalne knjižnice u ulozi pouzdanih digitalnih arhiva

Nacionalne su knjižnice vodeće kulturne ustanove čija je glavna uloga i zadatak da pohranjuju pismenu baštinu svoga naroda za buduće generacije. To im daje organizacijsku, stratešku i poslovnu pouzdanost. Sve nacionalne knjižnice neka su vrsta javnog zavoda jer djeluju u interesu naroda, a financiraju se iz državnog proračuna, pri čemu možemo govoriti o njihovoj financijskoj pouzdanosti. Njihov stručni kadar čine knjižničari i informacijski stručnjaci, koji su obrazovani i osposobljeni za organizaciju i sistematizaciju pohranjenoga znanja i informacija. Znaju kako prikupljati, pohranjivati i omogućiti pristup pohranjenim fondovima. Publikacije na novim medijima novi su oblik građe za koju je nacionalna knjižnica *a priori* zadužena. Svi postupci pri arhiviranju i omogućavanju pristupa elektroničkoj građi, kao i kod ostale tradicionalne građe, temelje na međunarodnoj praksi i standardima. To znači da su svi organizacijski, financijski, poslovni i proceduralni preduvjeti ispunjeni do te mjere do koje bi ih ispunila malo koja druga institucija bez odgovarajućih i potrebnih financijskih i kadrovskih sredstava. Pored toga, institucija obveznoga primjerka daje nacionalnim knjižnicama mogućnost prikupljanja cjelovite pismene produkcije u zemlji. Najveći je problem kod nacionalnih knjižnica arhivski informacijsko-tehnološki sustav, koji bi morao biti prilagođen sustavu OAIS. Postupci se mogu izvoditi ručno, međutim kod veće količine podataka potrebni su složeniji sustavi. Neke vlade

³⁷ An audit checklist for the certification of trusted digital repositories : draft for public comment. Mountain View, CA : RLG-OCLC, Aug. 2005. URL: <http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf> (15.12.2006.)

³⁸ DINI-Certificate. Document and Publication Repositories. Göttingen : Deutsche Initiative für Netzwerkinformation, Oct. 2003. URL: <http://www.dini.de/documents/Zertifikat-en.pdf> (7.2.2004.)

³⁹ Cijena dobivanja DINI-eva certifikata za članove neprofitnih organizacija je 50 €, a za članove profitnih organizacija 150 €. Cijena za one koji nisu članovi neprofitnih organizacija je 100 €, a za one koji nisu članovi profitnih organizacija 250 €.

dodatno financiraju razvoj takvih aplikacija, dok druge nacionalne knjižnice pokušavaju doći do potrebnih sredstava preko istraživačkih ili razvojnih projekata.

Osim nacionalnih, u svijetu naravno postoje i primjeri drugih knjižnica koje su preuzele ulogu pouzdanih digitalnih arhiva. Naprimjer sveučilišne knjižnice u Innsbrucku i Grazu, u Austriji, Sveučilišna knjižnica u Bratislavi, u Slovačkoj, ili Sveučilišna knjižnica u Brnu, u Češkoj, imaju ulogu depozitarnih regionalnih knjižnica i za svoje regije obavljaju funkciju pouzdanih digitalnih arhiva. Međutim, do izgradnje pouzdanih digitalnih arhiva došli su pomoću međunarodnih projekta jer su njihovi prioriteti vezane uz interese matičnih institucija, a to je omogućavanje pristupa "svježim" informacijama studentima i pedagoškim radnicima za potrebe pedagoškoga procesa i istraživanja. Sličan se problem pojavio kod specijaliziranih knjižnica, koje nude informacijsku potporu tekućem istraživačkom radu matičnih institucija. Školske i javne opće knjižnice nisu zainteresirane toliko za pohranu, koliko za omogućavanje pristupa različitoj građi, jer se bore za interese svojih osnivača.

Osim nacionalnih knjižnica, ulogu pouzdanoga digitalnoga arhiva za elektroničku dokumentarnu građu preuzimaju i nacionalne arhivske institucije ili specijalizirane nacionalne institucije za različitu digitalnu građu, primjerice RTV, glazbene nacionalne institucije i filharmonije, muzeji i slično. Svaka druga institucija, i ako bi uspjela uspostaviti optimalan arhivski sustav za pohranu elektroničke građe, ne bi mogla uspješno djelovati i obavljati tu funkciju bez ostalih spomenutih uvjeta. Zato su nacionalne knjižnice najprimjernije ustanove za dugoročnu pohranu pismene baštine.

3.4. Vrsta digitalnih repozitorija

Djelovanje digitalnoga arhiva, osim mnogih procesa vezanih za prikupljanje obradu, pohranu i pristup elektroničkoj građi, ovisi i o aplikaciji koja služi za smještanje i upravljanje tom građom u arhivu, t.j. digitalnom repozitoriju. Vodeća sila u razvoju i implementaciji digitalnih repozitorija je SAD. Većina danas upotrijebljenih aplikacija nastala je u okviru istraživanja na američkim sveučilištima, čiji je rezultat softver koji je u većini slučajeva dostupan na njihovim mrežnim

stranicama preko svjetske mreže. Među najpoznatijima su DSpace (www.dspace.org), koji su razvili u *Massachusetts Institute of Technology* (MIT) u suradnji sa *Hewlett-Packard Company*, i Fedora, koju su razvili u suradnji na Sveučilištima Cornell i Virginija.⁴⁰

DSpace⁴¹ je nastao kao aplikacija za dugoročnu pohranu tekstova, audio i video snimki, slika i baza podataka. Napisan je u Javi i djeluje na Linuxu, sustavu UNIX ili Windowsima XP. Koristi mrežno sučelje za pristup elektroničkim sadržajima. Omogućava različite načine unosa elektroničke građe u repozitorij zajedno s podacima u Dublinskom osnovnom skupu elemenata metapodataka (*Dublin Core*), automatski stvara kontrolni zbroj (*checksums*) za sve datoteke u repozitoriju, što pomaže u utvrđivanju njihove autentičnosti i prepoznavanju duplikata. Za složene datoteke koristi se standardom u XML-u METS (*Metadata Encoding Transmission Standard*), kojim održava veze između različitih komponenata elektroničke građe. Sustav automatski identificira format datoteka i veže ih na odgovarajući softver, koji omogućava njihov pristup. Taj se softver skladišti izvan DSpacea. Osim toga, sa DSpaceom treba posebno ugraditi relacijsku bazu, u koju se smještaju svi metapodatci. DSpaceom se koristi više od 150 institucija, većinom u sveučilišnim knjižnicama u cijelom svijetu.⁴²

Fedora⁴³ je kratica za *Flexible Extensible Digital Object and Repository Architecture*. Kao i DSpace, napisana je u Javi, djeluje sa sustavom Unix, Windows ili Mac i dostupna je kao *open source*. Aplikacija se može lako dograđivati, a može smjestiti oko 10 milijuna digitalnih objekata. Datoteke se mogu unositi u Fedora Object XML (FOXML) formatu ili u METS-u, što osigurava povezanost svih komponenata kod složenijih digitalnih objekata. Softver je još u razvoju. Fedora omogućava izvođenje svih procesa dugoročne pohrane tako što odvaja podatke od sučelja i aplikacija u arhitekturi sustava, čime olakšava migracije podataka. Veoma je fleksibilna i može se konfigurirati tako da izvodi sve procese i aktivnosti iz modela OAIS. Podržava i protokol OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*), što joj daje veću funkcionalnost kod internetskoga pretraživanja. Fedora je usto vrlo rasprostranjena, pogotovo u

⁴⁰ Falk, H. Digital archive developments. *The Electronic Library*. 21, 4(2003), str. 375-379.

⁴¹ Pennock, M. DSpace digital repository software, 12.6.2006. URL: <http://www.dcc.ac.uk/resource/technology-watch/dspace> (15.12.2006.)

⁴² Vidi popis korisnika DSpace na URL: <http://wiki.dspace.org/index.php/DspaceInstances> (15.12.2006.)

⁴³ Pennock, M. Fedora digital repository software, 24.11.2006. URL: <http://www.dcc.ac.uk/resource/technology-watch/fedora/> (15.12.2006.)

sveučilišnim knjižnicama. Imaju je i u Nacionalnoj knjižnici u Walesu, u Engleskoj, kao i u Nacionalnoj knjižnici Estonije.

Problem kod aplikacija koje su dostupne kao *open source* jest u tome da zahtijevaju mnogo znanja i sredstava za prevođenje i prilagođavanje već postojećoj infrastrukturi u knjižnici. Zato se za njih odlučuju knjižnice i institucije koje imaju posebne odjele s dovoljno informatičara koji se mogu posvetiti njihovom razvoju. Iskustva kolega estonske Nacionalne knjižnice, koji su mogli prilagoditi softver za repozitorij sredstvima europskoga projekta, pokazala su da su unatoč slobodnom pristupu softveru, bile potrebne najmanje dvije godine razvoja da bi dostigli normalnu razinu djelovanja. Pritom su njihovi informatičari morali više puta putovati u SAD na dodatno obrazovanje. Cijena koju su na kraju platili za sav razvoj je otprilike jednaka onoj koju bi platili da su nabavili neki od komercijalnih softvera. Zato su se austrijski kolege iz Nacionalne knjižnice u Beču na istom europskom projektu odlučili za komercijalnu varijantu. Već godinu dana ugrađuju aplikaciju DigiTool, koju za njih prilagođava izraelsko poduzeće ExLibris. DigiTool ima sve karakteristike koje imaju i drugi spomenuti softveri, s tom razlikom da je kompatibilan s integralnim knjižničnim sustavom ALEPH, koji proizvodi ista firma. Pošto se u Nacionalnoj knjižnici već koriste tim sustavom, nije bilo teško odlučiti se za nadgradnju.

3.5. Web arhivi

Repozitoriji digitalnih arhiva koje smo maloprije sagledali, upravljaju elektroničkom građom, t.j. digitalnim objektima, kao zasebnim cjelinama. Mogu preuzimati i složenije mrežne stranice i ugraditi ih u neki od formata u XML-u te tako održati njihovu originalnu strukturu, ali se pritom više puta mijenja njihov izgled. Osim toga, pristup građi moguć je samo preko metapodataka koji su smješteni u posebnoj relacijskoj bazi. Ti sustavi na neki način održavaju sličnu logiku kao i tradicionalna građa, s tom razlikom da su sposobni upravljati i najmanjim komponentama te građe. Mogu se koristiti za vrlo selektivno prikupljenu građu.

Međutim, kao što ćemo poslije vidjeti, mrežna građa sadrži drugačiju strukturu i njezine se stranice isprepliću s drugim mrežnim stranicama. Zato se arhitektura web arhiva razlikuje od

drugih digitalnih arhiva. Građa se u web arhivu mora pohraniti tako da simulira njezino postojanje u originalnom okruženju. To znači da osnova za prikupljanje, pretraživanje i pristup nisu umjetno stvoreni metapodatci, već URL-i koji se indeksiraju i smještaju u posebnu relacijsku bazu. Pretraživanje se obavlja pomoću mrežnoga sučelja i pomoću pretraživača, kao što je Google. Za razliku od internetskoga okruženja, ovdje postoje i mnoge verzije istih mrežnih stranica, stoga web arhivi moraju ponuditi i dodatnu mogućnost za vremensko sagledavanje tih stranica.

Zbog svega što smo istaknuli, za web arhiv ne možemo reći da je to samo zbirka mrežne građe, već je to složeniji sustav, koji uvelike ovisi o funkcionalnostima aplikacije za prikupljanje građe, o pristupu i metodama njezina prikupljanja te o pratećim aplikacijama koje omogućavaju cjelovito prikupljanje mrežne građe, njihovo indeksiranje i pravilno smještanje u samom repozitoriju. Osim toga, za pristup smještenoj građi potrebno je posebno sučelje s dodatnim funkcionalnostima, koje u običnim digitalnim arhivima ne nalazimo i koje se mogu razlikovati od arhiva do arhiva. Zbog njezinih specifičnih svojstava mnogo je teže aplicirati metode i strategije dugoročne pohrane mrežne građe u web arhivima. Nastoji se prenijeti praksa iz digitalnih arhiva, a za ocjenu primjerenosti tih postupaka još nije prošlo dovoljno vremena.

U nastavku su prikazana međunarodna iskustva u izgradnji web arhiva.

3.6 Pregled prakse i pristupa izgradnji web arhiva u svijetu

Iskustva u različitim državama pokazuju, da još uvijek ne postoji standardna praksa kod prikupljanja mrežne građe. Nacionalne su knjižnice svjesne potrebe prikupljanja i dugoročne pohrane publikacija na mreži, iako svaka od njih pritom pokušava pronaći primjerno rješenje s obzirom na raspoložive financijske i kadrovske resurse, kao i na postojeću infrastrukturu. Osim nacionalnih knjižnica, ulogu web arhiva preuzimaju i sveučilišne knjižnice, koje dijele funkciju depozitarnih organizacija s nacionalnim knjižnicama. Važan čimbenik za oblikovanje prakse prikupljanja mrežne građe jesu i mogućnosti koje daje Zakon o obveznom primjerku. Zakonski propisi na području obveznoga primjerka u većini se europskih zemalja mijenjaju u smjeru

uključivanja i mrežne građe. Svi spomenuti razlozi utječu na stvaranje različitih pristupa i strategija kod izgradnje web arhiva.⁴⁴

Prvi pokušaji prikupljanja mrežne građe izvedeni su oko 1996.-97. godine. Tada su se izgradnjom web arhiva počele baviti nacionalne knjižnice skandinavskih zemalja, neprofitno poduzeće u SAD-u *Internet Archive* i australska Nacionalna knjižnica. Slijedile su ih druge zemlje koje su s njima sudjelovale na istraživačkim projektima ili su gradile na njihovim znanjima i iskustvima.

U nastavku je kratak pregled međunarodne prakse prikupljanja mrežne građe i izgradnja web arhiva.

3.6.1. Početak izgradnje europskih web arhiva

Povijest prikupljanja mrežnih stranica u Europi započela je već 1991. godine s projektom **NORDINFO** (*Nordic Web Index*),⁴⁵ nordijskoga Savjeta za znanstvene informacije, čiji je cilj bio skupljati statističke informacije o URL-ima. U tu je svrhu izrađena aplikacija za pretraživanje Combineom, koju su zatim 1996. godine Šveđani dogradili u okviru projekta Švedske Kraljevske knjižnice **KULTURARW3** i stvorili aplikaciju za prikupljanje mrežne građe.^{46,47,48} Cilj projekta KULTURARW3 bio je istražiti metode za prikupljanje, arhiviranje i omogućavanje pristupa švedskim elektroničkim dokumentima koji su bili javno dostupni preko svjetske mreže. Od početka projekta 1997. godine robot za prikupljanje mrežne građe pokretao se u prosjeku svakih šest mjeseci. Prikupljanje traje u prosjeku dva mjeseca do pet mjeseci. Podatci su smješteni na magnetnim trakama. Sva građa prikupljena od 1997. do 2005. godine sadržavala je 305,85

⁴⁴ Christensen-Dalsgaard, B. ... [et al.]. Experiences and conclusions from a pilot study: web archiving of the district and country elections 2001. Final report for the pilot project "netarkivet.dk", February 2003. <http://netarchive.dk/publikationer/webark-final-rapport-2003.pdf> (15.12.2006.)

⁴⁵ NORDINFO (Nordic Council for Scientific Information). URL: <http://www.nordinfo.helsinki.fi/nordinfo/index.htm> (15.12.2006.).

⁴⁶ KULTURARW3. Stockholm : Royal Library (KB), Sweden's National Library. URL: <http://www.kb.se/kw3/ENG/Default.htm> (15.12.2006.).

⁴⁷ Arvidson, A. The collection of Swedish web pages at the Royal library - The web heritage of Sweden. // 68th IFLA Council and General Conference: Libraries for Life: Democracy, Diversity, Delivery, August 18th - 24th 2002, Glasgow, Scotland: [proceedings]. URL: <http://www.ifla.org/IV/ifla68/papers/111-163e.pdf> (15.12.2006.)

⁴⁸ Arvidson, A. The Kulturaw3 project - the Swedish Royal web archive // The Electronic Library. 16, 2(1998), str. 105-108. URL: <http://kulturaw3.kb.se/html/projectdescription.html>. (15.4.2003.)

milijuna datoteka (URL) ili 347.642 mrežnih mjesta, što je ukupno iznosilo 9.895 GB računalnoga prostora.⁴⁹

Do 2002. godine arhiv nije bio dostupan javnosti jer nije imao uređen pretraživač i dokumenti nisu bili dokumentirani ni bibliografsko obrađeni. Ako arhiv nije indeksiran, pretraživanje je moguće samo prema unaprijed definiranim URL-ima ili nadnevcima. Zbog zakonskih ograničenja web arhiv dostupan je samo u prostoru knjižnice.

Odmah nakon početka projekta KULTURARW3, na natječaju Evropske komisije odabran je za financiranje **projekt NEDLIB**⁵⁰ (siječanj 1998. - prosinac 2000.), čiji je cilj bio razviti osnovnu infrastrukturu i teorijski model sustava za pohranu elektroničkih publikacija te im omogućiti pristup i u budućnosti. U NEDLIB-u je sudjelovalo sedam europskih nacionalnih knjižnica: Norveške, Francuske, Njemačke, Italije, Nizozemske, Portugala i Švedske, Sveučilišne knjižnice u Helsinkiju i tri komercijalna nakladnika (dva iz Nizozemske i jedan iz Njemačke). Glavni koordinator projekta bila je nizozemska Kraljevska knjižnica.

U okviru projekta, *Center for Scientific Computing* (CSC) razvio je aplikaciju za prikupljanje mrežne građe, robot NEDLIB Harvester.⁵¹ Njegovo prvo testiranje izvedeno je u islandskoj Nacionalni knjižnici.⁵²

Godine 1997. u Švedskoj je osnovan forum za koordinaciju i izmjenu iskustava između različitih nacionalnih projekata u sjevernim zemljama. Na redovnim susretima odlučeno je da se projekt Kulturarw3 raširi i u druge nordijske zemlje te da se oblikuje mreža **NORDUnet2**. Godine 1998. nordijske nacionalne knjižnice odlučile su preuzeti odgovornost za razvoj nordijskoga web

⁴⁹ Kulturarw3 : Statistics. URL: <http://www.kb.se/kw3/Eng/Statistics.htm>. (15.12.2006.)

⁵⁰ Projekt NEDLIB: <http://nedlib.kb.nl/>. (15.12.2006.)

⁵¹ NEDLIB Harvester. Den Haag : Koninklijke Bibliotheek, [s.i.]. URL: <http://www.csc.fi/sovellus/nedlib/> (15.4.2006).

⁵² Hakala, J. Collecting and preserving the web : developing and testing the NEDLIB Harvester. // RLG DigiNews, 5, 2(2001). URL: <http://www.thames.rlg.org/preserv/diginews/diginews5-2.html> (15.4.2003), str. 2. Tom prilikom su skupili 5.750 domena (sve registrirane domene u Islandu u siječnju 2001.). Dobili su 565.169 dokumenata iz 1.426.371 URL-a, što znači u prosjeku oko pet na svakog Islandanina. Razlika između broja dokumenata i URL-a nastala je zbog velikog broja udvostručenih domena. Do udvostručavanja je došlo jer su mnogi poslužitelji imali po nekoliko imena i jer je više URL-a upućivalo na istu građu. Konačna varijanta islandskoga arhiva (komprimiranog 40-50% s GNU zip) obuhvaćala je 4.4 GB podataka.

arhiva. Tako je nastao projekt u kome je sudjelovalo pet nordijskih knjižnica (Finske, Norveške, Švedske, Danske i Islanda) **Nordic Web Archive** (NWA). Namijenili su oko 2.000.000 DKK (211.150 €) za razvoj aplikacije za pristup. To je bilo vrlo bitno jer su se u pet nordijskih zemalja već tada koristili dvama robotima, Combineom i NEDLIB Harvesterom. Nijedan od tih dvaju robota nije omogućavao pristup arhivu mrežnih stranica. Zato su članovi NWA odredili funkcionalne specifikacije za daljnji razvoj modula za prikupljanje mrežne građe i pristup arhivu, koji su nazvali *NWA Access Tool*.

U Norveškoj je već 1990. godine⁵³ stupio na snagu Zakon o obveznom primjerku, prema kome je Nacionalna knjižnica zadužena za pohranu dvaju primjeraka svih elektroničkih publikacija. Budući da je u to vrijeme Internet bio tek u početnom razvoju, sa sve većim objavljivanjem na tom mediju nastao je problem kako formirati praksu odabira, prikupljanja, opisa identifikacije i dugoročne pohrane publikacija na mreži. Kako bi našli odgovore na sva ta pitanja koja su se otvarala istovremeno s razvojem mrežnog arhiva, koji su po zakonu bili obvezni graditi i dopunjavati, u kolovoza 2001. godine započeli su **projekt PARADIGMA**.⁵⁴ Za razliku od ekstremnih selektivnih i automatskih pristupa prikupljanju publikacija na mreži, u ovom su projektu pokušali razviti sustav koji bi bio sposoban poluautomatski arhivirati mrežnu građu i koji bi omogućavao da se sačuvaju svi sastavni dijelovi i varijante dokumenata u cjelini. Metapodatci bi se kreirali automatski.

Za automatsko prikupljanje mrežne građe koristili su se robotom NEDLIB Harvesterom, a za selektivno sakupljanje HTTrackom, o kome će biti govora poslije. Za indeksiranje i pristup dokumentima u arhivu od 2002. godine rabili su *NWA Access Tool*. Prvo prikupljanje mreže odvijalo se u srpnju 2002. U kolovozu 2003. prikupili su 4.1 milijuna URL-a.⁵⁵

Finska Sveučilišna knjižnica u Helsinkiju imala je vrlo veliku ulogu u razvoju oruđa i metoda za prikupljanje i dugoročnu pohranu elektroničkih publikacija na mreži. Među većim projektima

⁵³ Zakon je bio prihvaćen 1989. godine, a tek slijedeće godine je stupio na snagu. Zato u literaturi nalazimo različita godišta.

⁵⁴ Albertsen, K. The Paradigma web harvesting environment. // 3rd Workshop on Web Archives, Trondheim, Norway, August 21st, 2003 : Proceedings / uredili Masanes, J. ; Rauber, A. ; Cobena, G. URL: <http://bibnum.bnf.fr/ECDL/2003/proceedings.php?f=albertsen> (15.12.2006.), str. 49-62.

⁵⁵ Koerbin, P. Report on the crawl and harvest of the whole Australian web domain undertaken during June and July 2005. Okt. 2005. URL: http://pandora.nla.gov.au/documents/domain_harvest_report_public.pdf (15.12.2006.)

valja spomenuti **projekt EVA**, koji je imao sljedeće ciljeve: odrediti kriterije odabira; uspostaviti načine za pridobivanje građe izravno od nakladnika; stvoriti mehanizme za prikupljanje takozvane »sive literature« na mreži; oblikovati bibliografski opis građe pomoću Finmarca, Dublinskog osnovnog skupa elemenata metapodataka (*Dublin Core*) i drugih metapodataka, uključivši preliminarne zapise; izgraditi web arhiv prikupljene građe i razviti softver koji je potreban za njegovo održavanje; strukturirati mogućnosti pretraživanja; proučiti mogućnosti dugoročne pohrane i riješiti pravna pitanja. Projekt EVA započeo je 1997. godine u suradnji sa Sveučilišnom knjižnicom u Helsinkiju, tri finska sveučilišta, nakladničkom kućom Edita i poduzećem *CSC Scientific Computing*. Prvo prikupljanje mreže uspješno su obavili u lipnju 2002. i od tada je redovito skupljaju.⁵⁶

U Finskoj su upotrijebili Linuxovu radnu stanicu koja je na jesen 2000. godine skupila oko 500.000 dokumenata. Finska je aktivno kao partner sudjelovala u projektu NEDLIB.⁵⁷ U okviru projekta EVA III, s financijskim sredstvima njihova ministarstva za kulturu, nastavili su s razvojem robota NEDLIB Harvester. Također su sudjelovali u projektu *Nordic Web Archive* u izgradnji sučelja za pristup mrežnom arhivu.

Godine 2002. objavljeno je prvo prikupljanje finske mreže s NEDLIB Harvesterom i tada su skupili 11,7 milijuna datoteka ukupne veličine 500 GB iz više od 40 milijuna lokacija. Na proljeće 2003. prikupili su 15.000 stranica koje su bile tematsko povezane s finskim parlamentarnim izborima. Na početku 2004. godine, nakon pet mjeseci skupljanja mreže, web arhiv sadržavao je osamnaest milijuna datoteka iz pedeset lokacija.

U proljeće 2004. godine zamijenili su robot NEDLIB Harvester s Heritrixom, u čijem su razvoju sudjelovali kao članovi konzorcija IIPC (*International Internet Preservation Consortium*). O projektu IIPC i Heritrixu bit će govora u posebnom poglavlju.

⁵⁶ <http://www.nla.gov.au/padi/topics/92.html#fin> (15.12.2006.)

⁵⁷ Hakala, J. Archiving the Web: European experiences. URL: <http://www.lib.helsinki.fi/tietolinja/0203/webarchive.html> (15.12.2006.)

3.6.2. Iskustva SAD-a u izgradnji web arhiva

U SAD-u postoje dva različita pristupa u izgradnji web arhiva. S jedne je strane web arhiv, koji na osnovi automatskoga domenskoga prikupljanja mrežne građe gradi *Internet Archive*, a s druge tematsko ili vrlo selektivno, koje obavlja Kongresna knjižnica (*Library of Congress*).

The Internet Archive je neprofitna organizacija iz San Francisca, čiji je glavni zadatak izgraditi javnu internetsku digitalnu knjižnicu. Osnovali su je Bruce Gilliat i Brewster Kahle u travnju 1996.⁵⁸ Građu im od 1996. godine sa šestomjesečnim vremenskim razmakom poklanja komercijalno poduzeće *Alexa Internet*. *Alexa Internet* je poduzeće za katalogizaciju mrežne građe, koje su također osnovali Brewster Kahle i Bruce Gilliat 1996. godine.⁵⁹ U to je vrijeme bilo vrlo teško dokazati važnost njihove aktivnosti. Počeli su s prikupljanjem mrežnih stranica predsjedničkih kandidata u SAD-u 1996. godine u suradnji sa *Smithsonian Institution* u Washingtonu. Nekoliko godina poslije arhiv je privukao veliku pažnju i tako su dobili nove zahtjeve za prikupljanje mrežnih stranica o izborima iz 2000. i 2002. godine. U studenom 1996. počeli su surađivati s *Alexom Internet*. To je neprofitno poduzeće za svoje potrebe prikupljalo svjetsku mrežu svakih osam tjedana po kriteriju najčešće posjećenosti mrežnim stranicama. Početne stranice skupljale su se po domenama, iako zbog ekstenzivnoga pristupa nisu uspjeli preuzeti sve promjene na webu. Dvomjesečno prikupljanje čini oko 10 TB podataka i više od 35 milijuna mrežnih mjesta. Do 2005. godine su u bazi *The Internet Archive* skupili oko 60 milijardi URL-a, odnosno oko 600 TB građe,⁶⁰ pa se veličina baze mjesečno povećava za oko 10 TB.

Do 1999. godine arhiv su čuvali na trakama pomoću ADIC 50, što je prouzrokovalo veliko kašnjenje kod pretraživanja i u pristupu. Osim toga, sretali su se s problemom autorskog prava,

⁵⁸ Kimpton, M. ; Ubois J. Year-by-Year: From an Archive on the Internet to an Archive on the Internet. // Web archiving / uredio Masanès, J. Berlin : Heidelberg ; Springer, 2006. Str. 202.

⁵⁹ Mohr, G., ... [et al.]. An introduction to Heritrix : an open source archival quality web crawler // 4th IAWA proceedings, 2004. URL: <http://1.1.1.1/330021196/336449912T070216081447.txt.binXMysM0dapplication/pdfXsysM0dhttp://www.iwaw.net/04/Mohr.pdf> (15.12.2006.)

⁶⁰ Stack, M. Full text search of web archive collections. // 5th International Web Archiving Workshop (IWA05), 22-23 September 2005, Vienna, Austria. URL: ; Kimpton, M. ; Ubois J. Year-by-Year: From an Archive on the Internet to an Archive on the Internet. // Web archiving / uredio Masanès, J. Berlin : Heidelberg ; Springer, 2006. str. 201-212 navodi da je u zahvaćenim podacima pedeset milijardi mrežnih stranica, trideset tisuća knjiga, trideset i šest tisuća zvučnih snimaka i petnaest tisuća filmova.

jer se mnogima nije svidjelo agresivno preuzimanje njihovih sadržaja. Rješenje se pojavilo kada je Alta Vista lansirala protokol za izuzimanje mrežnih stranica sa *robots.txt*.⁶¹

Godine 1999. godine je Andy Jewel za *Alexu Internet* napisao nov robot za prikupljanje mrežne građe, koji je bio sposoban filtrirati 16 milijardi i prikupiti oko 4 milijardi URL-a. Tada su prvi put upotrijebili datotečni format ARC za smještanje mrežnih stranica. ARC omogućava smještanje metapodatka zajedno s datotekama, kao i poveznicama između različitih datoteka. Robot kojim se prikuplja mrežna građa za *The Internet Archive* je vlasništvo *Alexe Internet*. *The Internet Archive* samo dobiva mrežnu građu iz *Alexe Internet* u svoj web arhiv i čuva je na tvrdim diskovima te indeksira njezin sadržaj.

Godine 2000. *The Internet Archive* započeo je suradnju s Kongresnom knjižnicom. Dvije godine poslije bio je jedan od osnivača konzorcija IIPC.

Web arhiv *The Internet Archive* slobodno je dostupan na mreži (www.archive.org), pristupa mu se pomoću sučelja i pretraživača *Wayback Machine* (WM), koji je završen u listopadu 2001. Za pretraživanje je potrebno znati točnu adresu URL-a, jer pretraživač ne može tražiti po sadržajima ni po metapodacima. Kad dobijemo traženu mrežnu stranicu, možemo pristupiti njezinim različitim verzijama u vremenu.

Nasuprot selektivnom pristupu, u Kongresnoj knjižnici odvija se projekt **MINERVA** (*Mapping the Internet: the Electronic Resources Virtual Archive*).⁶² To je web arhiv građen na tematskom principu.⁶³ URL-i koji bi mogli biti relevantni za određenu temu, događaj ili slično skupljaju se neprestano ili u određenim intervalima (naprimjer predsjednički izbori). U Kongresnoj knjižnici svjesni su činjenice da bi bilo puno jeftinije kad bi upotrijebili automatski pristup kod

⁶¹ Poslužitelj koristi **Robot Exclusion Standard** kako bi obišao sve web-poslužitelje koji su na njegove pretraživače prijavljeni i izvršio katalogiziranje (indeksiranje) njihovih stranica. Standard definira kako izraditi 'robots.txt' datoteku koja omogućava izabranim robotima da dohvate i indeksiraju sve sadržaje web-poslužitelja koji nisu zabranjeni prema nalogima u navedenoj datoteci. Koriste se ključne riječi 'User-agent:' i 'Disallow:' na način koji se dobije kao primjer u ispisu po kliku na ikonu 'robota' na dnu prve stranici ove knjige. Datoteka se smješta u korijenski imenik poslužitelja.

⁶² MINERVA (*Mapping the Internet Electronic Resources Virtual Archive*). Washington : Library of Congress. URL: <http://www.loc.gov/minerva/>. (15.4.2003.).

⁶³ Amerikanci koriste koncept web arhiva kao naziv za pojedinačne tematske zbirke. Tako postoje web arhiv izbora, web arhiv napada na New York itd.

prikupljanja građe.⁶⁴ Međutim, kod takvog pristupa nije moguće ocijeniti koje će mrežne stranice u budućnosti biti primjernije za istraživanja, iako očekuju veće korištenje arhiviranim sadržajima. Kod prikupljanja elektroničke građe surađuju s *The Internet Archive* i *Alexom Internet*.

Dvije veće tematske mrežne zbirke su Web arhiv napada na New York 11. rujna 2001. godine (5 TB podataka, više od 30.000 mrežnih mjesta i 332.000 URL-a) i Web arhiv izbora 2002. godine (1 TB podataka, više od 3.000 mrežnih mjesta i 82.000 URL-a).⁶⁵ Pronalaženje i katalogizaciju građe obavili su stručnjaci iz WebArchivist.org u formatu MODS (*Metadata Object Description Schema*), koji omogućava različit pristup samoj građi preko metapodataka.

Elektroničke dokumente u Kongresnoj knjižnici skupljaju uz pomoć robota sakupljača HTTrack. Za indeksiranje sadržaja i traženje po web arhivu koriste se licencnim pretraživačem Inktomi®.

U Kongresnoj knjižnici brinu se za dugoročnu pohranu pristupa elektroničkim publikacijama tako što sami odabiru strategiju za njihovu dugoročnu pohranu (migracija, emulacija ...).

Problematiku autorskoga prava sređuju tako što u MINERVI tijesno surađuju s biroom za autorska prava (*U.S. Copyright Office*). S njima pokušavaju definirati pravo za preuzimanje javno dostupne mrežne građe. Mnoge organizacije koje nude slobodan pristup informacijama na mrežnim stranicama žele da ih Kongresna knjižnica čuva za buduća istraživanja, iako Kongresna knjižnica nema eksplicitnih prava jer u zakonu (*Copyright Act*) ta ovlaštenja nisu određena. Autorima i vlasnicima autorskih prava šalju elektronsku poruku u kojoj ih obavještavaju da namjeravaju sačuvati njihovu mrežnu stranicu ili mrežno mjesto. Samo u slučaju da vlasnici autorskih prava ne žele da Kongresna knjižnica arhivira njihovu mrežnu stranicu, miču je iz web arhiva.

Kongresna knjižnica može odrediti i druge institucije koje se brinu za prikupljanje mrežnih stranica.

⁶⁴ Arms, W. Y. Collecting and Preserving the Web: The Minerva Prototype. // RLDigiNews, 5, 2(2001). URL: <http://www.rlg.org/preserv/diginews/diginews5-2.html#feature1>. (15.12.2006.)

⁶⁵ Schneider, S. M. ... [et al.]. Building thematic web collections : challenges and experiences from the September 11 web archive and the Election 2002 web archive. // 3rd Workshop on Web Archives, Trondheim, Norway, August 21st, 2003 : Proceedings / uredili Masanès, J. ; Rauber, A. ; Cobena, G. URL: <http://bibnum.bnf.fr/ecdl/2003/proceedings.php?f=schneider>. (15.12.2006.), str. 77-93.

3.6.3. Web arhiv Nacionalne knjižnice Australije

Australska Nacionalna knjižnica među prvima je u svijetu započela s praksom prikupljanja mrežne građe. Njihov je pristup donedavno bio isključivo selektivan. Međutim, selektivni pristup nije sasvim riješio pohranu cijele australske mreže jer do srpnja 2006., t.j. jedanaest godina poslije prvoga prikupljanja, uspjeli su skupiti tek 12,423 naslova mrežnih publikacija. S obzirom na kratak životni ciklus publikacija na mreži, znali su da taj pristup nije dovoljan. Zato zadnjih nekoliko godina također proučavaju mogućnosti uvođenja i automatskoga prikupljanja. U nastavku su opisana oba pristupa.

PANDORA je kratica za *Preserving and Accessing Networked Documentary Resources of Australia*, projekt koji se 1996. godine počeo izvoditi u australskoj Nacionalnoj knjižnici.⁶⁶ Cilj projekta bio je proučavanje strategije preuzimanja, katalogizacije, pohrane i omogućavanja pristupa svim relevantnim australskim publikacijama koje su dostupne na internetu, i na taj način izgradnja arhiva knjižnične elektroničke građe. Godine 2001. razvili su sustav za digitalno arhiviranje PANDAS (*PANDORA Digital Archiving System*)⁶⁷ koji je svojevrsan zajednički knjižnični sustav deset australskih knjižnica i omogućava integralno upravljanje elektroničkim publikacijama na mreži (t.j. njihov odabir, dobivanje autorskih prava, zahvaćanje datoteka, provjeravanje kvalitete, dodavanje metapodataka, administriranje itd.).⁶⁸ Osim toga, kupili su i

⁶⁶ U historiji PANDORE na str. URL: <http://pandora.nla.gov.au/historyachievements.html> (15.12.2006.) je zapisano, da su prva dva naslova odnosno mrežna mjesta prikupili listopada 1996.

⁶⁷ URL: <http://pandora.nla.gov.au/pandas.html>. (15.12.2006.)

⁶⁸ Prvi sustav koji je prethodio PANDAS-u i koji su razvili u samoj knjižnici tražio je puno rada u održavanju. Zbog sve veće baze podataka, sustav se morao neprestano razvijati, jer se već tada bližio gornjim granicama kapaciteta i tehničke potpore. Taj su problem pokušali riješiti 1998. godine u okviru projekta *Digital Services Project (DSP)*, koji bi im omogućavao integralnu tehničku infrastrukturu za sve digitalne usluge (Sustav za upravljanje digitalnim zbirkama). Nakon što je javni natječaj iz 1999. godine bio neuspješan, odlučili su se za modularni pristup, odnosno za razvoj manjih projekata koji bi osigurali odgovarajuća programska rješenja. Budući da u to vrijeme nije bilo odgovarajućeg sustava za digitalno arhiviranje, Australci su morali sami razvijati aplikaciju pa je u lipnju 2001. godine izašla prva verzija PANDORINOG Digitalnog Arhivskog Sustava (PANDAS). Druga, poboljšana verzija izašla je u kolovozu 2002. godine, a trenutačno vjerojatno koriste već treću. U okviru sustava za upravljanje procesima arhiviranja djeluje i repozitorij ili sustav za arhiviranje digitalnih objekata (*Digital Object Storage System (DOSS)*). Sustav je sastavljen iz poslužitelja E450, CLARION FC4700 i tračne jedinice StorageTek Tape Library. Godine 2002. knjižnica je potpisala ugovor s poduzećem InQuirion Pty Ltd. za računalnu aplikaciju TeraText, koja bi im trebala omogućiti distribuirano pretraživanje u bazi po metapodacima i po elektroničkoj građi, posebno slikovnoj građi te glazbenim snimkama.

PANDAS je bio sastavljen iz sljedeća četiri modula:

- Modul za prikazivanje, koji sadrži sučelje i odgovarajuće aplikacije za klijenta (*client*), omogućava vizualni pristup modulu aplikacijskoga sustava. Klijentsko sučelje dopušta pristup samo sučelju javnoga arhiva (PANDORA).
- Aplikacijski modul, koji sadrži specifične klijentsko/poslužiteljske (*client/server*) implementacije, omogućava funkcionalnosti i djelovanje sustava aplikacija.

sustav za arhiviranje digitalnih dokumenata DOSS (*Digital Object Storage System*), koji djeluje kao digitalni repozitorij. Proučili su različite mogućnosti smještanja datoteka mrežne građe, uključivši prijenos na poslužitelja u postojećem formatu te potpunu konverziju u »standardni format«. Nacionalna knjižnica oblikovala je nacionalnu strategiju za omogućavanje pristupa elektroničkim publikacijama na mreži. Veliku pažnju posvećuju problemu nepouzdanosti elektroničkih adresa, proučavaju mogućnosti za prevođenje standardnih brojeva u trajni identifikator URN, mogućnosti URI-a, odnosno PERL-a.

U kolovozu 2004. godine⁶⁹ web arhiv australske Nacionalne knjižnice sadržavao je 6,608 arhiviranih naslova mrežne građe (21.117.595 datoteka ili 702,1 GB). U godinu dana porastao je za 2.707 naslova (oko. 26.000.000 datoteka ili 925 GB), da bi u srpnju 2006. godine sadržavao 12,423 naslova mrežne građe (32.108.057 datoteka ili 1210,73 GB).⁷⁰ Neke dokumente u web arhivu tvori samo jedna datoteka (obično PDF) ili tisuću datoteka u različitim formatima, uključivši tekstove, zvučne snimke, slike ili video građu.

Prikupljanje mrežne građe izvodi se pomoću robota sakupljača HTTrack,⁷¹ koji na osnovi definiranih uputa zahvaća određene naslove. Prikupljaju se i elektronički dokumenti, koje

-
- Poslovni modul omogućava zajednički pristup različnim smještenim podacima te informacijsku infrastrukturu za aplikacijski modul.
 - Podatkovni modul sastavljen je od standardnih sučelja za pristup poslovnom modulu. Tu je relacijska baza u Oraclu (*Relational database Management*), podatkovni poslužitelji i WebDav dostupni datotečni poslužitelji.

Tehničku infrastrukturu PANDORE čini sljedeće:

- robot sakupljač mrežnih stranica i elektroničkih dokumenata (HTTrack),
- web poslužitelj (*Apache*),
- baza metapodataka (*Dynix system*),
- web sučelje, koje omogućava pristup arhiviranoj građi (*ALS web PAC*),
- hijerarhijski sustav datoteka za smještanje arhiviranih dokumenata (*UNIX files* na *IBN RISC 6000*),
- web sučelje za pristup UNIX-ovim datotekama, koje omogućava ograničene informacije o administriranju arhiva,
- sustav UNIX za smještanje i obnovu podataka (TIMBERLINE),
- LAN,
- mehanizmi za periodično izvještavanje.

⁶⁹ Koerbin, Paul: *The PANDORA Digital Archiving System (PANDAS): managing web archiving in Australia: a case study*. Paper presented at the 4th International Web Archiving Workshop on 16 September 2004 in Bath, UK.

⁷⁰ Informacije o statističkim podacima za 2006. godinu mogu se dobiti na URL: <http://pandora.nla.gov.au/statistics>. (15.12.2006.)

⁷¹ Prva prikupljanja interneta obavili su pomoću programa Harvest. To je aplikacija za indeksiranje koju je razvilo Sveučilište u Koloradu i prilagodili su je za australsku Nacionalnu knjižnicu, za potrebe arhiviranja interneta. Tada su se koristili rudimentarnim korisničkim sučeljem Pantrack i dodatno su razvili podsustav, koji je katalogizatorima omogućavao slanje i podešavanje zahtjeva za arhiviranje. Za administrativne metapodatke izradili su bazu podataka u Accessu, koja je djelovala neovisno od sustava. S pojavom sve većeg broja dokumenata u JavaScriptu, morali su zamijeniti postojeći alat sa WebZIP-om, a zatim i HTTrackom.

nakladnik po dogovoru šalje na CD-ROM-ima ili u komprimiranim datotekama preko interneta. Dokument se preslikava na poslužitelja i pomoću analitičkog programa (Linkbot) provjerava se nedostaje li u mrežnom mjestu koja datoteka. Provjeravanje se također izvodi ručno. Zatim se arhivirana verzija uspoređuje s elektroničkim originalom. Pokušava se dobiti verzija koja je najsljednija originalu. Također se nove verzije uspoređuju sa starima. Vanjske poveznice, koje nisu bile zahvaćene, onesposobljavaju se. Pritom se dodaju skripta tako da tko god klikne na poveznicu, dobije poruku da može prijeći u živo internetsko okruženje. Za svako novo mrežno mjesto pripremaju se naslovnice sa svim podacima i, ako je to potrebno, s programima za njihovo otvaranje. Sve te aktivnosti uzimaju sedam puta više vremena od obradbe tiskane građe.

Mrežna građa smješta se u tri oblika, a to su: arhivski original, koji je smješten u formatu TAR (*tape archive format*), original za pristup, koji sadrži promjene do kojih je došlo pri provjeravanju kvalitete, i metapodatci, također u formatu TAR, s informacijama o strukturi direktorija, imenima datoteka originalnoga mrežnoga mjesta i ostalo. Osim tih triju oblika, izrađuje se i kopija za prikazivanje, bez ikakvih kompresija, koja je dostupna na web poslužitelju i namijenjena je javnosti.⁷²

Za arhiviranje mrežnih dokumenata u australskoj Nacionalni knjižnici osnovana su dva odjela, Jedinica za građu digitalnoga arhiva i Jedinica za informacijsku tehnologiju. U prvoj je zaposleno dvadeset i četvero ljudi, a u drugoj sedam. Osim njih, arhiviranje elektroničkih publikacija obavljaju stručnjaci iz šest odjela.

Bibliografski opisi svih dokumenata u arhivu čuvaju se u katalozima svih deset knjižnica koje sudjeluju u mreži i u bazi Nacionalne bibliografije (zajednička baza 850 australskih knjižnica). Zbog zaštite komercijalnih interesa, neke publikacije imaju ograničenje pristupa do pet godina.

⁷² Više o tome u Koerbin, P. The PANDORA Digital Archiving System (PANDAS): managing web archiving in Australia: a case study. Paper presented at the 4th International Web Archiving Workshop on 16 September 2004 in Bath, UK. Cjeloviti model poslovnoga procesa opisan je u dokumentu Pandora Business Process Model, URL: <http://pandora.nla.gov.au/bmp.html> (15.12.2006.) (prvi put objavljen 1997. godine, obnovljen 2005. godine).

Dostupne su lokalno u prostorijama knjižnice. Naslove u arhivu indeksiraju komercijalni pretraživači, kao što su Google i Yahoo.⁷³

Za australsku Nacionalnu knjižnicu bitan je vrlo sistematičan pristup arhiviranju elektroničkih publikacija. Njihova pravila i smjernice za prikupljanje, obradbu i dugoročnu pohranu građe dostupni su svim zainteresiranim stručnjacima na njihovu mrežnom mjestu. Svaki je postupak detaljno opisan i može poslužiti kao praktičan uzor. Njihova su iskustva poslužila kao osnova za UNESCO-vu publikaciju *Guidelines for the preservation of digital heritage*.⁷⁴

Istovremeno s projektom PANDORA odvija se i projekt **PADI**, čiji je cilj prosljeđivanje informacija i omogućavanje pristupa većim međunarodnim izvorima informacija iz područja dugoročne pohrane elektroničkih publikacija.

Australci su svjesni nedostataka koje im donosi strog i selektivan pristup u izgradnji web arhiva. Kao prvo, djelatnici koji rade s isključivo novim informacijama moraju odlučiti jesu li određene publikacije na mreži važne za budućnost.⁷⁵ Sam rad je kadrovski i vremenski veoma zahtjevan i cijena svake sačuvane publikacije je zato veoma visoka.⁷⁶ Stoga su 2001. godine u australskoj Nacionalnoj knjižnici počeli istraživati izvodljivost automatskoga i manje selektivnoga pristupa.⁷⁷ Paul Koerbin⁷⁸ je 2005. godine objavio detaljni izvještaj o iskustvima domenskoga

⁷³ Phillips, M. E. Selective archiving of web resources: a study of acquisition costs at the National Library of Australia. // RLG *DigiNews*, June 15, 2005. URL: http://www.rlg.org/en/page.php?Page_ID=20666&Printable=1&Article_ID=1749. (15.12.2006.)

⁷⁴ Publikaciju *Guidelines for the preservation of digital heritage* je po narudžbi UNESCO-a pripremio Colin Webb, direktor Odjela za dugoročnu pohranu u Nacionalnoj knjižnici Australije u suradnji s drugim stručnjacima u svijetu. Publikacija je objavljena na stranicama UNESCO-a <http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>. (15.12.2006.)

⁷⁵ Phillips, M. E. Op. cit.

⁷⁶ Phillipsova u spomenutom djelu navodi da su izračunali da za obradu samo jedne mrežne stranice ili bibliografske jedinice na mreži trebaju ukupno 441 min. Cijena za njihovo arhiviranje je 68,36 AUD ili 105,23 € (1 AUD = oko 1,6 €). U usporedbi s arhiviranjem obveznoga primjerka knjiga (43,77 AUD ili 27,35€) ili obveznoga primjerka serijske publikacije (11,29 AUD odnosno 7€) više od pet puta je skuplje.

⁷⁷ Cathro, W. ; Webb, C. ; Whiting, J. Archiving the web : the PANDORA Archive at the National Library of Australia. A paper presented by Warwick Cathro at the Preserving the present for the future web archiving conference, Copenhagen, 18-19 June 2001. URL: <http://www.nla.gov.au/nla/staffoaoer/2001/cathro3.html> (15.12.2006.)

⁷⁸ Koerbin, P. u svome izvještaju iz 2005. godine (Report on the crawl and harvest of the whole Australian web domain... URL: http://pandora.nla.gov.au/documents/domain_harvest_report_public.pdf. (15.12.2006.) između ostalog navodi sljedeće nedostatke kod prikupljanja australske mreže:

- za provjeravanje kvalitete prikupljanja je bilo potrebno pregledati web arhiv;
- neke poveznice, koje su originalno bile dinamične stranice ili interaktivne baze podataka, nisu djelovale;

prikupljanja australske mreže, koji se odvijao u lipnju i srpnju 2005. godine. U tu svrhu unajmili su američku organizaciju *The Internet Archive*, koja im je u šest tjedana skupila 189,824,119 dokumenata ili datoteka u ukupnoj veličini 6.69 TB podataka iz 811,523 "hostova".

Ugovor sa *Internet Archive* uključivao je: prikupljanje mreže po australskoj nacionalnoj domeni tijekom četiri tjedna, indeksiranje, prijenos i instalaciju potrebnoga hardvera, *Middleware* i softvera, s odgovarajućim stručnjacima koji bi to izveli. Za pristup web arhivu su upotrijebili su sučelje *Wayback Machine*.

Cjeloviti preuzimanje mreže po australskim domenama bio je sedam puta veći od svih prikupljenih mrežnih stranica u arhivu PANDORA u zadnjih devet godina. Postigli su prilično veliku dubinu prikupljanja. Među nedostacima je navedena nesposobnost robota da identificira pomoćne datoteke i njihove skupove okvira (*frameset*). Usto, robot nije mogao preuzimati datoteke smještene na poslužiteljima sa pravilom *robots.txt* niti lokalizirati mrežne stranice izvan domene i geografske lokacije Australije, koje su važne za Australiju.

Problem je u tome što ne postoji metrički sustav u kojem bi se mogao utvrditi postotak preuzimanja cjelovitoga australskoga interneta. Jedino što se može učiniti jest da se usporedi popis registriranih domena sa zahvaćenim URL-ima.

-
- pristup web arhivu bio je uređen pomoću sučelja *Wayback Machine*; to sučelje omogućavalo je samo pristup određenom URL-u; isto tako koristio se metodom za prikazivanje stranica s pretraživačem koji upotrebljava JavaScript tako da su se prikazani sadržaji isprepletati sa živim mrežnim okruženjem;
 - u australskoj Nacionalnoj knjižnici analizirali su stranice tako da su usporedili kopije u web arhivu sa prikupljenim pomoću Googlea te usporedili kvalitetu i jednih i drugih; osim toga, provjerili su geografsku lokaciju stranica izvan domene *.au, kao i relevantnost stranica, koje su bile isključene sa *robots.txt*;

S obzirom na analizu dubine prikupljanja su našli nekoliko nedostataka:

- nepostojanje statistike datoteka u živom okruženju i u web arhivu, što je onemogućavalo njihovu usporedbu;
- zbog velike količine zahvaćenih stranica morali su se pouzdati na provjeru manjeg uzorka;
- problem su bile interne apsolutne poveznice u arhiviranim stranicama, koje nisu bile ponovno prepisane; kod njihova pregledavanja, vidjelo se da vode u živo okruženje, što daje utisak da je prikupljanje bilo uspješno;
- neučinkovitost i teškoće kod pretraživanja po arhiviranom sadržaju sa sučeljem *Wayback Machine*; datoteke su pojedinačno dostupne, iako je međusobna navigacija u određenim primjerima bila prekinuta;

Datoteke koje su stvarale teškoće i čije je prikupljanje bilo lakše s Heritrixom nego li s HTTrackom, jesu sljedeće:

- datoteke na *.html
- mrežna mjesta sa menijima u JavaScriptu
- mrežna mjesta napisana pomoću tehnologije *Flash* (v PANDORI koriste HEX editor za dostizanje funkcionalnosti)
- 444,214 URL-a poručivalo je odgovor s kodom 401, što znači da su to bile stranice koje zahtijevaju lozinku za dostup. Zanimljivo je da je taj broj mrežnih stranica činilo samo 0.234% zahvaćenih URL-a.

Drugo prikupljanje bilo je planirano za 2006. godinu. Još uvijek su vezani uz zakon o obveznom primjerku, koji isključuje mrežne stranice, pa ne mogu tu aktivnost uključiti u redovitu djelatnost, iako rade na promjeni zakonskih propisa.

3.6.4. Ostala bitna iskustva u izgradnji web arhiva

Projekti u sjevernim europskim zemljama, kao i projekt NEDLIB, uvelike su utjecali na nacionalne knjižnice ostalih europskih zemalja. Mnoge su počele s pokusnim prikupljanjem i izgradnjom web arhiva odmah nakon što su alati za prikupljanje bili javno dostupni na mreži. Internet Archive, kao i praksa australske Nacionalne knjižnice, također su imali veliki utjecaj u cijelom svijetu. Zanimljivo je da su i u Japanu i Kini također pokušali riješiti problem arhiviranja mrežne građe, a njihovi se rezultati mogu usporediti s praksom europskih zemalja.

Češki web arhiv

Moravska knjižnica u Brnu, koja ima status češke nacionalne knjižnice, 2000. je godine u suradnji s Nacionalnom knjižnicom u Pragu započela eksperimentirati s NEDLIB Harvesterom.⁷⁹ Suradivala je s Institutom za računarstvo Sveučilišta Masaryk, koji je pomogao s tehničkom i programskom potporom i istovremeno održavao češku verziju specifikacije Dublinskog osnovnog skupa elemenata metapodataka (*Dublin Core*). Robot im je prouzrokovao mnogo teškoća i zato su ga morali prilagoditi svojim potrebama. Za prikupljanje i stvaranje metapodataka upotrijebili su ili dogradili programe kao što su *Dublin Core Metadata Generator* i *URN:NBN unique identifier generator*. Analizirali su i mogućnost katalogizacije publikacija na mreži s Dublinskim osnovnim skupom elemenata metapodataka (*Dublin Core*).

Godine 2001. izveli su prvo pilotsko prikupljanje mreže po domeni *.cz. Zbog teškoća sa NEDLIB Harvesterom morali su prekinuti. Uspjeli su prikupiti ukupno 130 GB dokumenata. Nakon tog iskustva dodatno su razvili sučelje *Harvester Configurator* koje je omogućavalo bibliotekarima eksperimentiranje s tim alatom i samostalno namještanje parametara za traženje.

⁷⁹ Žabička, P. Archiving the Czech web: issues and challenges. // 3rd Workshop on Web Archives, Trondheim, Norway, August 21st, 2003 : Proceedings / uredili Masanès, J. ; Rauber, A. ; Cobena, G. URL: <http://bibnum.bnf.fr/ecdl/2003/proceedings.php?f=zabicka> (15.12.2006), str. 111-117.

U travnju 2002. izveli su drugo prikupljanje češkog interneta. U dva mjeseca uspjeli su prikupiti oko deset milijuna dokumenata (240 GB podataka) iz 30.000 domena s druge razine.

Kad je već bilo jasno da se NEDLIB Harvester neće više razvijati, počeli su pratiti aktivnosti *Internet Archiva* i IIPC. U 2005. godini upotrijebili su Heritrix, koji je bio slobodno dostupan kao sustav s otvorenim kodom. Sve to vrijeme usko su surađivali s Heritrixovim programerima. Za pristup web arhivu upotrijebili su sučelje IIPC, WERA (WEb aRchive Access) i NutchWax za indeksiranje. Ti moduli omogućavaju traženje po cijelom tekstu, u određenom vremenskom rasponu, po ključnim riječima ili URL-ima. Ne podupiru Boolove operatore ni kraćenje.

Javnosti je dostupan samo dio web arhiva, i to elektronička građa odabranih nakladnika s kojima je Knjižnica potpisala ugovor o autorskom pravu.⁸⁰ Ni njihov Zakon o obveznom primjerku ne uključuje mrežnu građu, zato su primorani za svaku elektroničku publikaciju na mreži posebno se dogovarati s vlasnicima autorskih prava.

Krajem 2005. godine njihov je arhiv bio veliki dva TB i sadržavao je 52 milijuna datoteka.

Web arhiv danske Kraljevske knjižnice

Danski Zakon o obveznom primjerku iz 1997. godine uključuje samo statične digitalne publikacije, iako je već tada vrijedilo pravilo da se mogu prikupljati na dobrovoljnoj osnovi. Godine 2003. njihov je digitalni arhiv već sadržavao 7.000 dokumenata.⁸¹ U svakom slučaju, nedostatak zakonske uredbe koja bi omogućila prikupljanje elektroničkih publikacija u okviru njihovih redovnih aktivnosti, prouzročio je danskoj Kraljevskoj knjižnici mnogo teškoća. To se posebno pokazalo 2001. godine, kada su počeli eksperimentirati s različitim metodama prikupljanja elektroničkih publikacija na mreži. Danska je Kraljevska knjižnica u studenom 2001., u okviru pilotskoga **projekta netarkivet.dk**, testirala različite metode odabira, prikupljanja i arhiviranja publikacija na mreži. Njihov je pristup bio tematski, tj. izbori koji su održani 20. studenoga 2001. Tu su temu odabrali jer su očekivali veliku produkciju na mreži i jer se odnosila na cijelu kraljevinu. Osim toga, taj je događaj bio vremenski ograničen, što je

⁸⁰ Češki WebArchiv: <http://www.webarchiv.cz/> (15.12.2006)

⁸¹ Christensen-Dalsgaard, B. Web archive activities in Denmark. // RLG DigiNews, 8, 3(2004). URL: http://www.rlg.org/en/page.php?Page_ID=17661#article0. (15.12.2006)

omogućavalo stvaranje zatvorene zbirke mrežnih stranica. Usto bi materijal mogao biti upotrijebljen za različita istraživanja i vrednovanja.⁸² Budući da njihov zakon o obveznom primjerku ne uključuje publikacije na mreži, morali su proučiti različite načine dobivanja autorskih prava.

Za pristup arhivu upotrijebili su sučelje *Nordic Web Archive*, koji se tada tek izgrađivao. Dugoročna pohrana nije bila planirana.

Za prikupljanje mreže koristili su se NEDLIB Harvesterom, WGET-om i RoboSuiteom. Prikupljanje je trajalo šest tjedana u odabranim općinama i pokrajinama. S NEDLIB Harvesterom zahvaćali su odabrane mrežne stranice jednom mjesečno, tjedno i dnevno. S WGET-om su zahvaćali mrežnu građu svakodnevno i svakog sata, a s RoboSuiteom, samo jednom dnevno. Za svaku mrežnu stranicu ili mrežno mjesto morali su dobiti dopuštenje autora.

Godine 2005., poslije prihvaćanja novoga zakona o obveznom primjerku, koji je uključivao i danski web, su u Kraljevskoj knjižnici započeli s iscrpnim prikupljanjem mrežne građe.^{83,84}

Pristup Nacionalne knjižnice Francuske u izgradnji web arhiva

Nacionalna knjižnica Francuske je u suradnji s istraživačkim institutom INRIA razvila program za prikupljanje mrežnih stranica, Xyleme.⁸⁵ Kod prvih pokusa prikupljanja mrežne građe njihov je pristup obuhvaćao dvije faze:

- prikupljanje građe po domenama,

⁸² Christensen-Dalsgaard, B. ... [et al.]. Experiences and conclusions from a pilot study: web archiving of the district and country elections 2001. Final report for the pilot project "netarkivet.dk", February 2003. <http://netarchive.dk/publikationer/webark-final-rapport-2003.pdf> (15.12.2006.)

⁸³ Pritom su se koristili trima različitim strategijama:

- Prikupljanje svih mrežnih stranica s relevantnim domenama četiri puta godišnje (*Cross-section harvesting*) - s Heritrixom.
- Selektivno češće (svakodnevno) skupljanje mrežnih stranica u okviru 80-tih domena - s HTTrackom.
- Tematsko prikupljanje mrežnih stranica dva do tri puta godišnje - s HTTrackom.

⁸⁴ Andersen, B. The DK-domain: in words and figures. URL: http://1.1.1.1/269697092/270240792T060816100652.txt.binXMysM0dapplication/pdfXsysM0dhttp://netarchive.dk/publikationer/DFreyv_english.pdf (15.12.2006.)

⁸⁵ Abiteboul, S. [e tal.]: A first experience in archiving the French web. // Research and advanced technology for digital libraries : 6th European Conference, ECDL 2002, Rome, Italy, September 16-18, 2002 : Proceedings / uredili Agosti, M. ; Thanos, C. Berlin ... etc. : Springer, 2002. Str. 1-15.

- određivanje prioriteta unutar same domene s obzirom na česte promjene mrežnoga mjesta i njegove važnosti. Njegovu važnost određivali su pomoću broja ulaznih poveznica (*in-going links*)⁸⁶ na to mrežno mjesto. Tu takozvanu *page-ranking* metodu već primjenjuju najpoznatiji svjetski sustavi za pretraživanje, kao što je Google.

Pomoću robota prikupljača Xyleme i nekoliko običnih osobnih računala su 2001. godine u prosjeku su svakodnevno zahvaćali do nekoliko milijuna mrežnih stranica. Tada su predviđali da s domenom na *.fr postoji oko milijarda mrežnih stranica, odnosno dvadeset milijuna URL-a, što čini tek jedan do dva posto cijele svjetske mreže. Nakon toga su knjižničari odlučivali o tome koje će se stranice sačuvati, a potom su utvrđivali koje je stranice robot zaobišao te koje bi bilo potrebno potražiti ili dopuniti.

Da bi povećali kapacitete skladištenja, koristili su se kompresijskom tehnikom, a sve promjene su jasno prikazane u shemi XML, na osnovi:

- trajnog identifikatora (*Persistent identifier*) mrežnih stranica koji tvore pomoću URL-a i nadnevka preuzimanja svakog dokumenta (URL, nadnevak),
- informacija o kopijama mrežnih mjesta i njihovih obnovljenim verzijama,
- potpore za pretraživanje u vremenu i za listanje po web arhivu.

Zatim su, 2002. godine, predstavnici francuske Nacionalne knjižnice Francuske započeli pregovore s *Internet Archive*. Zajedno s drugim europskim partnerima 2004. godine osnovali su IIPC i u okviru koalicije započeli s razvojem novog alata za skupljanje mreže, Heritrixa.

Islandski web arhiv

Osamdeset posto stanovnika Islanda ima pristup internetu kod kuće, na poslu, u školama i knjižnicama. Zato je internet vrlo bitan element njihove kulture i društva.

Za dugoročnu pohranu publikacija na mreži zadužena je Nacionalna i sveučilišna knjižnica. Pritom usko surađuje s ostalim nordijskim zemljama, također preko projekata, kao što su

⁸⁶ Razlikujemo dvije vrste poveznica među mrežnim stranicama: ulazna poveznica (*in-link*) i vanjska ili izlazna poveznica (*out-link*). Ulazne poveznice dovode nas do određene mrežne stranice, izlazne ili vanjske poveznice s te stranice na drugu.

NEDLIB, NWA, i programa NORDUnet. Godine 2001. u okviru projekta NEDLIB testirali su robot za prikupljanje mreže i prikupiti islandski web u okviru domena na *.is. Njihov je web arhiv 2005. godine sadržavao oko 200 GB podataka.⁸⁷ Pristup mrežnom arhivu omogućava im sučelje koje su nordijske zemlje razvile u okviru projekta *Nordic Web Archive*.

Mrežne stranice u domeni *.is prikupljaju se tri puta godišnje, a od svibnja 2005. zahvaćaju kontinuirano odabrana mrežna mjesta, koja obrađuju društveno političke teme i diskusije. Nacionalna i sveučilišna knjižnica ima zakonsku obavezu preuzimanja sve elektroničke građe, a od 1. 1. 2003. također i one na internetu. Mrežni je arhiv slobodno dostupan javnosti.

Islandska Nacionalna i sveučilišna knjižnica također je članica konzorcija IIPC.

Iskustva Nacionalne knjižnice Njemačke

U Nacionalnoj knjižnici u Frankfurtu 2000. i 2001. godine odvijali su se prvi eksperimenti automatskoga prikupljanja mreže. Budući da nisu postigli željenu kvalitetu, odlučili su se na to da zbirku građe selektivno. Nakladnici dobrovoljno donose građu u arhiv Deposit.DBB.DE, koji je samo lokalno dostupan. U 2006. godini nastavljaju s tom praksom, iako već imaju zakonske osnove za prikupljanje cijele njemačkoga interneta.

Njihova zbirka sadrži elektroničke publikacije, koje već nekoliko godina skupljaju na različite načine. U cjelini skupljaju doktorske disertacije,⁸⁸ radi čega su sklopili ugovore sa sveučilištima, i druge elektroničke publikacije, koje im po dogovoru šalju nakladnici.

Zanimljivo je i to da su u nacionalnoj knjižnici odlučili da ne specijaliziraju obradu elektroničke građe, jer je to samo novi oblik publiciranja koji će u budućnosti biti sve češća. Taj zadatak obavljaju katalogizatori koji imaju više predznanja iz područja informacijske tehnologije, jer elektroničku građu treba znati prepoznati i treba pratiti sve postupke koji su vezani uz djelovanje poslužitelja te o tome obavještavati nakladnike.

⁸⁷ Podatci dobiveni u okviru projekta Web Cultural Heritage i objavljeni na str. URL: <http://www.webarchiv.cz/culture-2000-documents/>. (15.12.2006.)

⁸⁸ METADISS. Frankfurt : Die Deutsche Bibliothek. URL: <http://deposit.ddb.de/metadiss.htm> (17.4.2003.)

Izgradnja web arhiva mrežne građe Demokratske Republike Kine

U Demokratskoj Republici Kini od nastanka interneta broj mrežnih stranica svake godine eksponencijalno raste.^{89,90} Početkom 2001. godine na Pekinškom su sveučilištu započeli **projekt WebInfoMall** (<http://www.infomall.cn/>), čiji je cilj bio istražiti mogućnosti i metode dugoročne pohrane mrežne građe. U tu su svrhu upotrijebili robot Tianwang i odgovarajuće formate za smještanje mrežnih stranica. Poslužili su se distribuiranom arhitekturom za prikupljanje mreže.⁹¹ Aplikacija je bila sposobna preuzeti i smjestiti oko milijun stranica na dan te upravljati s oko milijardu stranica (150 TB).

Pomoću robota Tianwang do lipnja 2004. godine prikupili su 0.7 milijarde mrežnih stranica (10,6 TB), s oko 5 TB elektroničke građe pored mrežnih stranica. U studenom 2005. godine u okviru projekta skupili su još jednu milijardu mrežnih stranica (15 TB podataka).

Web arhiv koji sadrži kinesku građu na mreži od 2001. je godine slobodno dostupan javnosti. Upotrebljavaju ga za različite interdisciplinarne studije i istraživanja iz područja humanističkih znanosti, kao i tehnologije. Dodatno su razvili različite aplikacije koje omogućavaju makropanoramsku analizu kineskoga interneta u vremenskoj i prostornoj dimenziji,⁹² naprimjer model za traženje mrežnih informacija na specifičnim područjima ili lokalne djelatnosti na mreži u provinciji Kwangtung.

Da se zbog političke cenzure ne bi izgubili važni aspekti kineskoga političkoga, društvenoga i historijskoga trenutka, Institut za kineske studije u Heidelbergu, u Njemačkoj, u suradnji sa Sinološkim institutom Sveučilišta u Leidenu, u Nizozemskoj, na poticaj prof. Rogera Wagnera

⁸⁹ Yan, H. ... [et al.]: A new data storage and service model of China web InfoMall. // 4th International Web Archiving Workshop, 2004. URL: www.iwaw.net/04/proceedings/Hongfei.pdf. (15.12.2006)

⁹⁰ Na osnovi publikacije 14th Statistical Survey Report on the Internet Development in China. China Internet Network Information Center (CNNIC), 20 July 2004. URL: <http://www.cnnic.net.cn/html/Dir/2004/07/22/2406.html> (15.12.2006) u Kini su 1994. uveli Internet i njihove prve mrežne stranice su nastale 1995. godine. Krajem 2003. godine je bilo već 595,550 registriranih mrežnih mjesta sa domenom *.cn, a krajem lipnja 2004. 6,27 milijuna, što znači da je bio prirast od 32,2% u godinu dana.

⁹¹ Lian'en, H. ; Hongfei, Y. ; Xiaoming, L. Engineering of Web InfoMall: The Chinese Web Archive // World Engineers' Convention 2004, November 2-6, 2004, Shanghai, China. URL: <http://net.pku.edu.cn/~yhf/refpaper/hle2004/1.jpg - 6.jpg>. (15.12.2006)

⁹² Na osnovi zadnjeg vremenskog redoslijeda promjena mrežnih stranica su izračunali frekventnost promjena i prosječan životni ciklus mreže. Ustanovili su da se 50% svih mrežnih stranica obnavlja u intervalima manjim od dva mjeseca. Polovinu svih komercijalnih mrežnih mjesta doživi manje od dva mjeseca, dok vladina ili obrazovna mrežna mjesta imaju dva puta veći životni ciklus.

započeo je **projekt DACHS** (*Digital Archive for Chinese Studies*, <http://www.sino.uni-heidelberg.de/dachs/>), kao dio Europskog centra za digitalnu građu na području kineskih studija.^{93,94} Projekt su započeli na proljeće 2001. godine. Njegov je cilj bio arhiviranje onoga dijela interneta koji odražava socijalna i politička događanja u vezi s Kinom.

U svibnju 2005. godine DACHS je sadržavao 1.970.555 datoteka (36,277 GB). Među njima, u 262.939 datoteka (5,8 GB) bile su znanstvene revije i članci na raznim jezicima, u prvom redu engleskom i kineskom, 235.449 datoteka (4,3 GB) činila je elektronička građa pojedinaca i različiti dokumenti o Kini na raznim jezicima, a najveći dio, 1.265.857 datoteka (24 GB) sadržavao je mrežna mjesta i stranice iz BBS i drugih mrežnih mjesta. Oko 979 datoteka (0,15 GB) dobili su na račun donacija, 982 datoteke (0,427 GB) činila je filmska građa, a ostalo su bili razni diskusijski forumi (204.349 datoteka ili 1,6 GB).⁹⁵

Arhiv DACHS gradi se dobrovoljnim priložima pojedinaca iz cijeloga svijeta i sustavnim pretraživanjem i prikupljanjem građe iz svjetske mreže. Prednost je u tom što pojedinci, koji pridonose arhivu, dobro poznaju određena područja i sami osvježuju informacije i stranice. Problem su vanjski pokušaji njihova prekrajanja, dugoročne pohrane arhiva te osiguravanje sredstava za održavanje samog sustava (kadrovskih i informacijsko-tehnoloških).

Digitalni arhivi mrežnih publikacija Hrvatske

U Hrvatskoj je započeto s praksom skupljanja mrežne građe već 2002. godine, kada je 19. veljače hrvatski Ured za ISBN i ISMN potpisao Ugovor o predaji elektroničkih publikacija Nacionalnoj i sveučilišnoj knjižnici u Zagrebu (NSK), sa deset važnijih elektroničkih nakladnika.⁹⁶ Na proljeće te iste godine izrađeni su pravilnici i obrasci s uputom za predaju dobrovoljnoga primjerka NSK.

⁹³ Wagner, R. G. ; Gross, J. Harvesting the web, preserving Chinese voices. The Digital Archive for Chinese Studies (DACHS), 11 Nov. 2004. URL: <http://www.sino.uni-heidelberg.de/dachs/publications/taipei2004.rtf>. (15.12.2006)

⁹⁴ Gross, J. Learning by doing : the Digital Archive for Chienes Studies (DACHS). // 3rd Workshop on Web Archives, Trondheim, Norway, August 21st, 2003 : Proceedings / uredili Masanès, J. ; Rauber, A. Cobena, G. URL: <http://bibnum.bnf.fr/ecdl/2003/proceedings.php?f=gross>. (15.12.2006), str. 103-110.

⁹⁵ Lecher, H. E. Small scale academic web archiving : DACHS. // Web archiving / uredio Masanès, J. Berlin Heidelberg : Springer, 2006. Str. 215.

⁹⁶ Živković, D. Hrvatski ured za ISBN i ISMN : izveštaj o radu 1.1.-14.11.2002.

Godinu dana poslije NSK je u zajedničkoj suradnji sa Srcem razvila arhivski sustav, koji temelji na Digitalnom web arhivu DAMP (*Digital Archive for Web Publications*),^{97,98} koji je povezan s katalogom NSK CROLIST. DAMP je sastavljen iz šest različitih modula, koji omogućavaju prikupljanje elektroničke građe na mreži, njihov opis, smještanje u arhiv te pristup. Osnovni metapodatci o građi automatski se prenose u CROLIST. Sustav se može lako namjestiti s obzirom na učestalost i dubinu prikupljanja određenih mrežnih stranica. Usto je moguće i odrediti koje dijelove na mrežnom mjestu i koje tipove podataka žele prikupiti.

Datoteke koje pripadaju jednom mrežnom mjestu ili publikaciji čuvaju se u odvojenim direktorijima zajedno s damp.xml datotekom koja se automatski stvara kod konačnog prikupljanja. Damp.xml datoteka sadrži kompletan popis sve prikupljene građe na jednom mrežnom mjestu. Osnovni metapodatci koji se stvaraju kod opisa građe sastoje se od imena datoteke, URL-a, vremena preuzimanja, tipa datoteke, kontrolnog zbroja (*checksum*) datoteke, koda HTTP statusa i tipa sadržaja *HTTP headera*.

Prikupljanje je vrlo selektivan i po naredbi, počevši od definiranog URL-a. Elektronička se građa obrađuje slično kao i tradicionalna.

Drugi velik projekt na području arhiviranja mrežne građe u Hrvatskoj je POOPD - Prikupljanje, odabir, obrada i pohrana dokumenata sa službenih mrežnih stranica državnih tijela Republike Hrvatske.⁹⁹ Projekt je započet u svibnju 2004. godine u suradnji Hrvatske informacijsko-dokumentacijske referalne agencije (HIDRA) i Sveučilišnog računskog centra, Sveučilišta u Zagrebu (Srce). Pritom se koriste iskustvima NSK-a. Službene izvore informacija državnih tijela Republike Hrvatske na mreži prikupljaju automatski pomoću robota i tu građu smještaju u arhiv mrežnih dokumenata AMD. Elektronički dokumenti pohranjuju se u digitalnom arhivu DAMIR (Digitalni Arhiv Mrežnih IzvoRa). Omogućen je mrežni pristup građi, a moguće je i pretraživanje punog teksta sadržaja. Do rujna 2006. godine prikupili su 72.790 dokumentata, od čega su

⁹⁷ Milinović, M. ; Topolščak, N. The architecture of DAMP : a system for harvesting and archivnig web publications. // Widwisawn 3, 3(2005). URL: http://widwisawn.cdlr.strath.ac.uk/Issues/Vol3/issue3_3_1.html. (15.12.2006.)

⁹⁸ Willer, M. ; Milinović, M. DAMP: sustav za preuzimanje i arhiviranje obveznog primjerka hrvatskih mrežnih publikacija. // Tehnični in vsebinski problemi klasičnega in elektronskega arhiviranja : zbornik referatov dopolnilnega izobraževanja s področij arhivistike, dokumentalistike in informatike. 4(2005). Str. 491-500.

⁹⁹ Više o projektu POOPD na mrežnom mjestu URL: <http://www.hidra.hr/hidra/hidra.htm>. (19.2.2007.)

katalogizirali 5.002 dokumenta. Radi bolje obrade, na mrežnim su stranicama objavili i preporuke za autore i nakladnike.¹⁰⁰

Web arhiv nizozemske Kraljevske knjižnice

Nizozemska nema zakonskih propisa o obveznom primjerku, međutim njihova je suradnja s domaćim nakladnicima toliko razvijena da nakladnici sami dobrovoljno šalju primjerke svojih publikacija nizozemskoj Kraljevskoj knjižnici (*Koninklijke Bibliotheek* - KB).

Godine 1995. u KB-u počeli su s izgradnjom digitalnoga repozitorija DNEP (*Depot van Nederlandse Electronische Publicaties*) koji je bio namijenjen dugoročnoj pohrani dobivene elektroničke građe. U prosincu 2002. IBM je završio izgradnju digitalnoga repozitorija DIAS (*Digital Information Archiving System*) na temelju modela OAIS, koji je postao osnova digitalnoga arhiva e-Depot.¹⁰¹ E-Depot je tako postao formalni arhiv najvećih nizozemskih nakladničkih kuća, Kluwer i Elsevier Science, te drugih članova Nizozemskog društva nakladnika (DPA). KB trenutačno pregovara sa stranim nakladničkim kućama, kao što su BioMed Central, Blackwell Publishing, Oxford University Press i Taylor&Francis Group. Kod izrade e-Depota Nizozemci su iskoristili iskustva vezana uz dugoročnu pohranu elektroničke građe iz projekta NEDLIB, čiji su bili glavni koordinator, i pokušali da ih primijene u praksi.

Pomoću robota NEDLIB Harvester obavljeno je više pokušaja automatskoga zahvaćanja nizozemskoga interneta u okviru domene *.nl. Godine 2002. objavljena je i studija Hansa Verhoevena o izvedivosti dugoročne pohrane mrežnih stranica u nizozemskom e-Depotu.¹⁰²

¹⁰⁰ Rezultate projekta opisuje Horvat, T. E-arhiv službenih publikacija i dokumenata tijela javne vlasti RH. // 35. skupština Hrvatskoga knjižničarskog društva, Plitvička jezera, Hotel Jezero 27.-30.9.2006. URL: <http://www.hidra.hr/hidra/10godina/plitvice.pdf>. (19.2.2007.). Više informacija o sustavu AMD može se naći u Garvas Delić, A. ; Milinović, M. Prikupljanje, odabir, obrada i pohrana dokumenata sa službenih mrežnih stranica tijela javne vlasti Republike Hrvatske. // 2. Kongres hrvatskih arhivista, Dubrovnik, listopad 2005. godine / Hrvatsko arhivističko društvo ; Državni arhiv u Dubrovniku. URL: http://www.hidra.hr/hidra/10godina/POOPD_za_HAD.pdf. (19.2.2007.)

¹⁰¹ Jansen, H. Permanent access : the e-Depot at the Koninklijke Bibliotheek. URL: <http://www.nla.gov.au/webarchiving/JansenHans.rtf> (15.12.2006.)

¹⁰² Verhoeven, I. H. Archiving web publications. Amsterdam : IBM Netherlands, Dec. 2002. (IBM/KB Long-Term Preservation Study Report Series Number 6). URL: http://www.kb.nl/hrd/dd/dd_onderzoek/reports/6-webpublications.pdf. (15.12.2006.), p.45.

Digitalni arhiv Nacionalne knjižnice Novog Zelanda

Novozelandska Nacionalna knjižnica tijesno surađuje s australskom Nacionalnom knjižnicom u traženju rješenja za dugoročnu pohranu nacionalne mrežne građe. Za razliku od Australaca, Novozelandčani se ne opterećuju autorskim pravima, jer ih njihov zakon o obveznom primjerku obvezuje da skupljaju i publikacije na internetu.

Od Australaca su preuzeli tehnologiju PANDAS i program za prikupljanje mreže HTTrack. Godine 2002. započeli su sa selektivnim prikupljanjem novozelandskih elektroničkih publikacija na mreži. Od prosinca 2002. godine prikupili su više od 400 publikacija i cjelovitih mrežnih mjesta.

Znatno su pridonijeli teoriji i praksi dugoročne pohrane elektroničke građe bogatim metapodatkovnim modelom.¹⁰³

Web arhiv japanske Nacionalne kongresne knjižnice

S projektom *Web Archiving Project (WARP)*^{104,105,106,107} japanska je Nacionalna kongresna knjižnica (*National Diet Library - NDL*) 2002. godine započela pokus selektivnoga prikupljanja mrežne građe. U svoj web arhiv pohranjuju mrežna mjesta o centralnoj vladi, o gradovima koji su nekad bili veliki i stabilni, a s vremenom su nestali, te stranice koje obrađuju razne velike događaje. Rezultati projekta uzet će se u obzir pri izradi novog zakona o obveznom primjerku.

¹⁰³ Više informacija o novozelandskoj praksi možemo naći u sljedećim izvorima: PADI - Web archiving: URL: <http://www.nla.gov.au/padi/topics/92.html#NZ>. (15.12.2006); National Library to Capture New Zealand's Digital Heritage URL: <http://www.natlib.govt.nz/bin/media/pr?item=1085885702>. (15.12.2006); National Library of New Zealand (Te Puna Matauranga o Aotearoa) Act 2003. URL: <http://www.natlib.govt.nz/files/Act03-19.pdf>. (15.12.2006); National Library to lead electronic harvesting URL: <http://www.natlib.govt.nz/bin/media/pr?item=1064531843> (15.12.2006); National Library of New Zealand - Digital Library Development Review URL: http://www.natlib.govt.nz/files/ross_report.pdf (15.12.2006); Collection Policy / National Library of New Zealand: URL: <http://www.natlib.govt.nz/en/about/1keypolcollections.html> (15.12.2006)

¹⁰⁴ Web Archiving Project (WARP), URL: <http://warp.ndl.go.jp/> (15.12.2006)

¹⁰⁵ Prototype Evaluation: Web Archiving Project (WARP). Appendix 15. URL: <http://www.crl.edu/PDF/PCWAappendices1-20.pdf> (14.10.2005.)

¹⁰⁶ Bibliotheca Universalis – Collections – NDL. URL: http://www.kb.nl/gabriel/bibliotheca-universalis/en/ndl/Bibliotheca%20Universalis_ndl.htm (14.10.2005.)

¹⁰⁷ Kurahashi, T. Collection development of the National Diet Library by In: CDNLAO Newsletter, No. 53, July 2005. URL: <http://www.ndl.go.jp/en/publication/cdnla0/053/534.html>. (14.10.2005.)

Projekt WARP zaključen je 2005. godine. Za preuzimanje mrežnih stranica koriste se aplikacijom Web Robot ili Wget 1.5.3. U listopadu 2003. godine web arhiv sadržavao je već 2.439.000 datoteka na 927 adresa. Arhiv je s dopuštenjem vlasnika autorskih prava dostupan javnosti.

Dok je prioritet projekta WARP prikupljanje površinske mreže, drugi projekt, **DNAVI** (*Database Navigation Service*), nastoji sačuvati dubinsku mrežu. DNAVI je portal koji je presnimio veliku količinu informacija iz različitih japanskih mrežnih mjesta i omogućava korisnicima navigaciju po raznim bazama podataka.

NDL istražuje mogućnosti automatskoga ekstenzivnoga prikupljanja mreže po domeni *.jp zajedno sa svim tehničkim i operativnim izazovima, koji bi taj pristup tražio.

Za opis elektroničke građe, kao osnovni metapodatkovni format služi im Dublinski osnovni skup elemenata metapodataka (*Dublin Core*), koji sadrži dodatne kvalifikatore, što omogućava njihov prijenos u japanski format MARC.

NDL pokušava odrediti standarde kao što su uvjeti i intervali prikupljanja mrežne građe. Ustanovili su da bi bilo optimalno kad bi mrežna mjesta preuzimali jednom mjesečno, a mrežne revije ovisno o učestalosti njihovih promjena. Mrežne stranice, za koje nemaju dopuštenje vlasnika autorskih prava, dostupne su samo u prostorijama knjižnice.

Web arhiv Velike Britanije

Bitnu ulogu kod gradnje web arhiva u Velikoj Britaniji ima konzorcij UKWAC (*UK Web Archiving Consortium*).¹⁰⁸ Čini ga šest članova koji su vrlo zainteresirani za dugoročnu pohranu publikacija iz svog područja na webu, t.j.: *British Library*, *Nacionalni Arhivi*, *Joint Information Systems Committee* (JISC), Nacionalne knjižnice Škotske i Walesa te *Wellcome Library*. Njihov pristup je vrlo selektivan i očigledno je da je uzor australska Nacionalna knjižnica. Ne samo da koriste aplikaciju za skupljanje HTTrack, već i njihov sustav za upravljanje sadržajima u arhivu PANDAS.

¹⁰⁸ Bailey, S. ; Thompson, D. IKWAC : Building the UK's first public web archive. // D-Lib Magazine, 12. 1(2006). URL: <http://dlib.org/dlib/january06/thompson/01thompson.html>. (15.12.2006.)

U ožujku 2005. godine mrežni su arhiv otvorili za javnost.¹⁰⁹ Pretraživanje arhiva omogućeno je preko hijerarhijske klasifikacije tematskih odrednica (*scheme of topic headings*). Mogućnosti pretraživanja veoma su slične kao kod pretraživača Google ili Yahoo. Arhiv se koristi sustavom za pretraživanje s aplikacijom u otvorenom kodu (*open source*) Lucene, koji je sposoban tražiti i po sadržajima pojedinačnih mrežnih stranica unutar samoga web arhiva.

U konzorciju su propisana veoma stroga pravila u vezi sa prikupljanjem mrežnih stranica i načinom njihove obrade. Svaka članica konzorcija može unaprijed obraditi novu mrežnu publikaciju u bazi. Pritom postaje odgovorna za njezino slijeđenje, dugoročnu pohranu i obnavljanje u arhivu. Velika se pažnja posvećuje stjecanju autorskih prava za dugoročnu pohranu i omogućavanje javnog pristupa mrežnoj publikaciji.

Kao i u Australiji, sustav PANDAS je centraliziran repozitorij i baza metapodataka. Metapodatci su smješteni i u lokalnim katalozima. Lokalizacija i pristup elektroničkoj građi mogući su pomoću trajnih identifikatora preko centralnoga ili lokalnih kataloga.

Posebna se pažnja posvećuje dugoročnoj pohrani sadržaja u web arhivu. Najviše teškoća prouzrokuju im tehničke strane sustava za arhiviranje. PANDAS nije fleksibilan, teško se prilagođava novim situacijama. Usto zahtijeva dodatne stručnjake i mnogo znanja iz informacijske tehnologije, koja je neophodna za analizu i prikupljanje cijelih mrežnih stranica.

Britanci također surađuju s međunarodnim konzorcijem IIPC u razvoju specifikacija i zahtjeva nove generacije alata za skupljanje mreže.

Uz spomenut konzorcij UKWAC, u Velikoj Britaniji postoje i drugi poticaji, naprimjer BBC (*British Broadcasting Corporation*) koji je po zakonu zadužen za dugoročnu pohranu cijele svoje produkcije. I BBC posvećuje veliku pažnju dugoročnoj pohrani produkcije na mreži, zbog čega i grade web arhiv.¹¹⁰

¹⁰⁹ Mrežni arhiv je dostupan na URL: <http://www.webarchive.org.uk>. (15.12.2006.)

¹¹⁰ Smith, C. Building an Internet Archive System for the British Broadcasting Corporation. // *Library Trends*, 54, 1(2005), str. 16-32.

3.6.5. Web arhivi kao rezultat akademskih istraživanja

Osim spomenutih iskustava u nacionalnim knjižnicama, postoje i razni eksperimenti na sveučilištima u prikupljanju mrežne građe. Međutim, u akademskim institucijama nije glavni cilj pohrana mrežne građe, a još manje njihovo dugoročno čuvanje, već eksperimentiranje s njihovim sadržajima.

Takvo istraživanje obavili su u Sveučilištu u Lisabonu, gdje su 2003. godine skupili oko 3.8 miliona URL-a ili 78 GB podataka,¹¹¹ i to pomoću robota sakupljača *tumba!*¹¹² Njihov je cilj bio karakterizacija portugalske mreže.

Drugi primjer nalazimo u suradnji Odjela za informatiku Ekonomskoga fakulteta u Ateni i Litavskoga sveučilišta u Vilnusu. Prikupljenu mrežnu građu upotrijebili su za sadržajnu analizu pomoću *text-mining* metoda.¹¹³ Sustav kojim su se koristili za izgradnju web arhiva bio je sastavljen iz sljedećih komponenata: robot sakupljač, sustav za upravljanje sadržajima (*content manager*) i modul za grupiranje (*clustering module*).

Još jedan primjer prikupljanja mrežne građe u okviru internetskoga istraživanja nalazimo u Austriji. Austrijski Zakon o obveznom primjerku ne uključuje mrežnu građu. Niti amandmani koje je Parlament prihvatio u srpnju 2000. godine nisu uključili mrežne publikacije. Zato ih njihova nacionalna knjižnica ne uključuje u svoj fond. Neovisno o tom su 1999. godine Odjel za softversku tehnologiju na Tehnološkom sveučilištu u Beču i austrijska Nacionalna knjižnica u okviru projekta AOLAP (*Austrian Online Archive Processing*)¹¹⁴ pokušali prikupiti austrijske

¹¹¹ Koerbin, P. Report on the crawl and harvest of the whole Australian web domain undertaken during June and July 2005, Okt. 2005. URL: http://pandora.nla.gov.au/documents/domain_harvest_report_public.pdf. (15.12.2006.)

¹¹² Gomes, D. ; Silva, M. J. A characterization of the Portuguese web. // 3rd Workshop on Web Archives, Trondheim, Norway, August 21st, 2003 : Proceedings / uredili Masanès, J. ; Rauber, A. ; Cobena, Gregory, 2003. Str. 63-76. URL: <http://bibnum.bnf.fr/ecdl/2003/index.html>. (15.12.2006.)

¹¹³ Lampos, C., [et al.] : Archiving the Greek web. In: 4th International Web Archiving Workshop on 16 September 2004 in Bath, UK, 2004. URL: <http://1.1.1.1/281285604/336449912T070216081135.txt.binXMysM0dapplication/pdfXsysM0dhttp://www.iwaw.net/04/Lampos.pdf> (15.12.2006.)

¹¹⁴ Rauber, A. ; Aschenbrenner, A. ; Witvowt, O. Austrian online archive processing : analyzing archives of the world wide web. // Research and advanced technology for digital libraries : 6th European Conference, ECDL 2002, Rome, Italy, September 16-31, 2002 : Proceedings / uredili Agosti, M ; Thanos, C. Berlin ... etc. : Springer, 2002. Str. 1-15.

publikacije na mreži. U lipnju 2002. godine su eksperimentirali s NEDLIB Harvesterom, a u svibnju 2001. godine s Combineom. Tako su stvoriti web arhiv od 150 GB, koji je bio sastavljen od 2,7 milijuna mrežnih stranica ili 21.000 mrežnih mjesta. Rauber, Aschenbrenner i Witvowt su 2002. godine objavili rezultate istraživanja o lokacijama i distribuciji mrežnih poslužitelja u Austriji, kao i o distribuciji formata datoteka na raznim poslužiteljima.¹¹⁵ Sadržaj arhiva je još uvijek predmet različitih tehnoloških, kulturnih i socioloških analiza.

* * *

Opisana iskustva pokazuju vrlo različita rješenja i pristupe u izgradnji web arhiva.¹¹⁶ S jedne strane, skandinavske zemlje i *Internet Archive* odlučili su se za automatsko prikupljanje mrežne građe. Njihova strategija sastoji se iz snimki stanja (*snapshot*) u jednakomjernim intervalima određenog broja dostupnih publikacija na mreži. Druge nacionalne knjižnice, kao što su australska ili kanadska, obrađuju publikacije na mreži jednako kao i tradicionalne publikacije i njihov je pristup vrlo selektivan. Obično se proces prikupljanja građe odvija ručno. Između automatskoga i ručnoga prikupljanja jest tematsko prikupljanje u kome pokušavaju u potpunosti obuhvatiti sve što je publicirano na mreži u okviru određene teme ili događaja. Takve se zbirke građe u Kongresnoj knjižnici u SAD-u, Britanskoj knjižnici, australskoj Nacionalnoj knjižnici, kao i u drugim knjižnicama. Nijedna od tih metoda sama po sebi nije optimalna. Branitelji strogo selektivnog prikupljanja građe svjesni su da je njihov pristup mnogo skuplji i da pritom gube većinu publikacija, koje zbog kratkog životnog ciklusa ne dočekaju da budu odabrane i smještanje u web arhiv. Osim toga, vrlo je teško za njih reći hoće li biti zanimljive budućim korisnicima. Na drugoj je strani automatsko prikupljanje građe kojim se ne rješava problematika dubinske mreže i prikupljanje publikacija koje su stvorene u složenim formatima. Lokalizacija građe u tako stvorenim web arhivima vrlo je zapletena jer nema potpunoga bibliografskoga opisa.

Upravo iz svih spomenutih razloga sve se više nacionalnih knjižnica odlučuje za kombinirani pristup. Čak i australska Nacionalna knjižnica, koja je slovila kao najveća tvrđava ručnoga i

¹¹⁵ Rauber, A., Aschenbrenner, A., and Witvowt, O. ... Op. cit.

¹¹⁶ Christensen-Dalsgaard, B. ... [et al.]. Experiences and conclusions from a pilot study: web archiving of the district and country elections 2001. Final report for the pilot project "netarkivet.dk", February 2003. <http://netarchive.dk/publikationer/webark-final-rapport-2003.pdf> (15.12.2006.)

selektivnoga pristupa, popustila je i zadnje dvije godine eksperimentira s automatskim prikupljanjem.

Problem koji ostaje otvoren jest dugoročna pohrana građe u web arhivima koji se grade automatskim prikupljanjem. U novije vrijeme konzorcij IIPC pokušava naći rješenje tog problema definirajući postupke i procese pri prikupljanju mreže, a svojim se aktivnostima zauzima za standardizaciju tih postupaka i upotrijebljenih formata. Usto su članovi spomenutoga konzorcija pridonijeli razvoju web arhiva i s izradom novih alata za prikupljanje mrežne građe, koji omogućavaju kvalitetnije prikupljanje i smještaj te građe u web arhivu. O djelovanju IIPC-a bit će govora na drugom mjestu.

3.7 Situacija u Sloveniji

NUK pokušava uspostaviti web arhiv još od 2002. godine. Pomoću različitih projekata o kojima će biti govora u posebnom poglavlju, NUK je pokušao pridobiti sredstva i znanje za razvoj takvog arhiva. Tako je 2002. godine započeo suradnju s istraživačima IJS-a koji su bili zaduženi za razvoj alata za prikupljanje, arhiviranje i pristup mrežnoj građi. Ti su alati zajedno sa web arhivom morali biti gotovi već krajem 2004. godine, iako se zbog različitih razloga to nije dogodilo ni do kraja 2006. godine. Dodatan je problem bio taj što u Sloveniji nije bilo druge institucije koja bi mogla pomoći pri takvu razvoju u NUK-u,¹¹⁷ a s druge je strane NUK funkcionirao samo sa sistemskim administratorima, bez odjela i bez stručnjaka s područja informatike koji bi mogli preuzeti razvoj web arhiva. Godine 2006. počele su se provoditi organizacijske promjene i tada su dva informatičara započela s eksperimentiranjem novim alatima koje je razvijao konzorcij IIPC.

¹¹⁷ U 44. i 45. čl. slovenskog Zakona o knjižničarstvu definirani su zadaci knjižničnoga informacijskoga servisa. Informacijsku i tehničku potporu morao bi pružati Institut informacijskih znanosti u Mariboru (IZUM). Međutim, po njihovoj interpretaciji zadaci IZUM-a su ograničeni na bibliografski skupni katalog i njegovo djelovanje, a ne i na razvoj digitalne knjižnice, kako se po suvremenoj interpretaciji i očekuje. NUK je oduvijek bio ovisan o tehničkom razvoju IZUM-a, sam nije osnovao svoju računalnu jedinicu sve do reorganizacije 2005. godine, kad je bilo jasno da ovakvo stanje više ne zadovoljava potrebe NUK-a.

Inače NUK zajedno s ostalim knjižnicama (292) u skupnom katalogu Slovenije (COBISS) obrađuje elektroničke publikacije na materijalnim medijima od početka njihova nastanka.¹¹⁸ Publikacije koje su dostupne *online* također se obrađuju, iako su se naslovi stranih serijskih publikacija koje je nabavljao konzorcij ljubljanskog Sveučilišta posebno obrađivale bez njihova prijenosa u digitalni arhiv. Trenutačno se mogu naći naslovi čija poveznica više ne djeluje jer nisu više dostupni.

Digitalnu knjižnicu Slovenije NUK je osnovao 2005. godine i ona trenutačno sadrži digitalizirane zbirke koje su mrežno dostupne. Među njima su članci suvremenih znanstvenih¹¹⁹ i starijih¹²⁰ revija, fotografije, razglednice i crteži poznatih pjesnika, pisaca, znanstvenika, muzičara, političara i likovnih umjetnika, slovenski zemljovid od 1548. do 1924. godine, nacrti Ljubljane od početka 19. do sredine 20. stoljeća, ilustrirani plakati iz 19. stoljeća itd. U portalu dLib.si mogu se naći i digitalizirane snimke starih muzičkih kompozicija od početka 20. stoljeća sve do Drugoga svjetskoga rata. Zbirka i dalje iz dana u dan raste. Planira se uključivanje digitaliziranih kopija svih slovenskih novina od 1991. godine nadalje, kao i akademskih istraživačkih radova.

Sva spomenuta građa obrađena je i dostupna preko COBISS-a ili preko portala dLib.si. U pozadini još nema digitalnoga arhiva u pravom smislu. NUK još uvijek razvija strategiju dugoročne pohrane i s novim projektima planira uvođenje repozitorija sa svim modulima, koji bi omogućili automatizaciju procesa dugoročne pohrane, koji se zasad obavlja ručno.

¹¹⁸ Godinu dana nakon uvođenja COBISS-a u Sloveniji, 1998. godine bilo je moguće katalogizirati mrežne publikacije i omogućiti njihov pristup preko URL-a. Taj je proces je bio pojednostavljen 1999. godine, s uvođenjem alata CORC (projekt OCLC) u IZUM-u. Tako su među prvima bile obrađene elektroničke publikacije dobivene preko različitih konzorcija.

¹¹⁹ Trenutačno su digitalizirane revije: *Psihološka obzorja* od 2000. godine, *Documenta praehistorica* od 2001-2004. godine, *Jezik in slovstvo* od 2003.-2004. godine, geografske periodične publikacije pod naslovom *Dela* od 1999-2004. U tijeku je nacionalni projekt digitalizacije svih znanstvenih djela koja su bila financirana iz javnih sredstava.

¹²⁰ Među starijim naslovima serijskih publikacija koje su u cjelini dostupne preko portala dLib.si jesu: *Ljubljanski zvon* (1881-1941), *Novi akordi* (1901-1914) i *Nova muzika* (1928-1929).

4. ALATI I METODE PRIKUPLJANJA I POHRANE MREŽNE GRAĐE

4.1. Softverske aplikacije za prikupljanje mrežne građe

U engleskom jeziku postoji više izraza za aplikacije za prikupljanje mrežne građe. Nazivaju ih paukom (*spider*), gmizavcem (*crawler*), žetecem (*harvester*), robotom ili skupljačem (*gatherer*). Svaki od tih različitih termina asocira na način na koji prikupljaju građu na mreži.

Roboti djeluju na osnovi predefinirane skupine parametara.¹²¹ U aplikaciji možemo odrediti početan broj URL-a (*seed URL list*) i zatim robot preko njihovih poveznica prelazi na niže razine i tako prikuplja sve stranice koje odgovaraju definiranim kriterijima.

Prvi roboti bili su izrađeni sredinom devedesetih godina da bi omogućili stvaranje mrežnih indeksa. Među njima je tada bio najpoznatiji pretraživač Alta Vista. Međutim, problem kod tih pretraživača bio je u tome da nakon indeksiranja i nalaženja određenih mrežnih stranica nisu mogli zadržati i pohraniti te stranice.

U okviru projekta Kulturaw3,¹²² čiji je cilj bio proučiti načine automatskog prikupljanja mrežnih stranica, prvi je put rođena ideja da se upotrijebi tehnologija tadašnjih pretraživača te da ih se dogradi tako da budu sposobni zadržati i pohraniti građu. Tako je nastao i Combine, prvi robot za prikupljanje mrežne građe.

U *The Internet Archive* ustanovili su da postoje četiri pristupa pri prikupljanju mreže:¹²³

- prikupljanje u širinu (*broad crawling*)
- usredotočeno prikupljanje (*focused crawling*)

¹²¹ Hakala, J. Archiving the web : European experiences. // Program: electronic library and information systems, 38, 3(2004), str. 176-183.

¹²² KULTURARW3. Stockholm : Royal Library (KB), Sweden's National Library. URL: <http://www.kb.se/kw3/ENG/Default.htm> (15.4.2003.).

¹²³ Sigurdsson, K. Incremental crawling with Heritrix. // 5th International Web Archiving Workshop (IWAW05), 22-23 September 2005, Vienna, Austria. URL: <http://www.iwaw.net/05/papers/iwaw05-sigurdsson.pdf>. (15.12.2006.)

- kontinuirano prikupljanje
- eksperimentalno prikupljanje

Prikupljanje u širinu (*broad crawling*) i **usredotočeno prikupljanje** (*focused crawling*) veoma su slični, oba pristupa temelje na strategiji stvaranja snimaka (*snapshot*) i periodičnoga prikupljanja. Prikupljanje je ponovljivo, iako uvijek započinje iz definiranih ulaznih URL-a. Informacije o prethodnim prikupljanjima ne koriste se, osim u slučaju promjena kod nekih namještanja kako bi se izbjegle pogreške. Kada je elektronička građa preuzeta, robot mora smjestiti identifikaciju (*fingerprint*) URI-a i pojednostaviti strukture podataka kako bi ih mogao održati u nekom redoslijedu dok ih ne procesira.

Pri **kontinuiranom prikupljanju** robot više puta ekstenzivno ili intenzivno obiđe jednu te istu građu. Da bi spriječio stvaranje duplikata, mora imati određen kontrolni mehanizam koji mu omogućava da izluči već posjećene mrežne stranice. Pritom se koristi takozvanom "rastućom strategijom" (*incremental strategy*), koja mu omogućava pamćenje svih promjena u zahvaćenim stranicama. Svakako je to mnogo teže dostići ako je veći broj ciljnih mrežnih stranica. Robot mora računati i na vjerojatnost promjena i s obzirom na njih prilagoditi prikupljanje. Prisutan je i problem formata. Datoteke u HTML-u doživljaju više i češće promjene nego li slike ili grafičke datoteke. Također se mnogo rjeđe mijenjaju datoteke u formatima Word ili PDF. Što je veća datoteka, očekujemo manje promjena. Kod većine mrežnih mjesta najviše se mijenja naslovnica. Stranice koje ne sadrže metapodatke privremene su ili se češće mijenjaju. Sve to je potrebno uzeti u obzir prilikom namještanja robota.

Pri prikupljanju sa "rastućom strategijom" često se događa da između dva preuzimanja dolazi do toliko promjena na mrežnim stranicama da roboti ne uspiju sve prikupiti. Heritrix se koristi upravo tom strategijom i to je problem koji pokušavaju riješiti u njegovu razvoju.

Eksperimentalno prikupljanje koristi se za testiranje različitih tehnika prikupljanja s obzirom na vrste sadržaja, redoslijeda, različitih protokola i analiza, odnosno za arhiviranje rezultata prikupljanja.

U nastavku su predstavljeni različiti roboti koji su pomogli u razvoju prakse prikupljanja mrežne građe.

Combine

Robot Combine¹²⁴ razvilo je poduzeće NetLab iz Lunda po uzoru na već postojeće sustave za pretraživanje, kao što su Yahoo i Altavista. Sučelje za pristup Zebra razvilo je dansko poduzeće Index Data. Prva verzija Combinea bila je namijenjena samo skupljanju informacija o URL-ima i nije zadržavala, odnosno prikupljala dokumente na mreži. Poslije su nordijski partneri NORDINFO odlučili prilagoditi postojeći robot tako da prikuplja i pohranjuje mrežne stranice. U okviru projekta Kulturarw3 Šveđani su u suradnji s NetLabom dogradili Combine i prilagodili ga za arhiviranje.

Combine je izrađen u programskom jeziku PERL5 i C++ te djeluje na poslužitelju SUN Solaris. Danas ga još uvijek koriste u Švedskoj, istovremeno s drugim alatima za prikupljanje mreže.

NEDLIB Harvester

NEDLIB Harvester je proizvod CSC-a (*The Center for Scientific Computing*) iz Finske. Izrađen je u programskom jeziku C. Prva verzija završena je u siječnju 2000., a zadnja verzija 1.2 izašla je u listopadu 2001. godine.¹²⁵ Modul za indeksiranje razvijen je poslije. U razvoju NEDLIB-a upotrijebljene su postojeće aplikacije, kao što su programi za računanje kontrolnog zbroja (*checksum*) MD5 (*Message Digest*) koji je potreban za automatsko stvaranje identifikatora za datoteke, koje su pohranjene na poslužitelju, i za kontrolu duplikata. Koristi platformu Linuxa odnosno bazu MySQL 3.23 (ili novije)¹²⁶ gdje pohranjuje popise URL-a. NEDLIB Harvester djeluje na vrlo skromnoj opremi.¹²⁷

Kao i većina robota, NEDLIB Harvester započinje prikupljanje iz određene mrežne stranice (URL) i prati poveznice dok ne prikupi sve do zadnje stranice. Djeluje tako da automatski stvara

¹²⁴ COMBINE. Lund: NetLab, 1999, URL: <http://www.lub.lu.se/combine/> (15.4.2003.).

¹²⁵ NEDLIB Harvester. <http://www.csc.fi/sovellus/nedlib/>. (15.12.2006.)

¹²⁶ Starije verzije operacijskog sustava Linux (MySQL 3.22 i starije verzije) su bile primjerne samo za testiranje jer su mogle prenijeti datoteke do najviše dva GB veličine. Zato su preporučili upotrebu novije verzije 3.23 koja je mogla prenijeti datoteke jednake veličine samog sustava.

¹²⁷ Sveučilišna knjižnica u Helsinkiju je ožujka 2001. godine kupila poslužitelj Sun E450 sa procesorom 480 MHz CPU, jednim gigabajtom memorije i kapacitetom diska 8 x 36.4 Gb. Skromna oprema je bila dovoljna za prva prikupljanja finških mrežnih stranica pod uvjetom da su datoteke bile arhivirane na trakovima.

"kataloge" ili skupove povezanih datoteka za svaki dan prikupljanja u svojoj datotečnoj strukturi. Unutar tih kataloga otvaraju se podređeni katalozi s najviše 2000 datoteka. Kad dostigne taj broj, otvori se novi katalog. NEDLIB ne smješta sve datoteke iz istoga mrežnoga mjesta u jedan katalog, već ih raspršuje po raznim katalozima. Datoteke dobivaju kod MD5 koji se stvara na osnovi URL-a i sadržaja datoteke. Za svaki dokument otvara se datoteka s metapodacima u kojoj se skupljaju podatci o URL-ima i njihovi tehnički podatci. To znači da je teško locirati pojedinačne URL-e u arhivu bez datoteke sa metapodacima. Također je teško izbrisati sve datoteke iz određenoga mrežnoga mjesta u slučaju da ih vlasnik želi ukinuti. Teško je utvrditi koliko je stranica iz jednoga mrežnoga mjesta već bilo prikupljeno.

Na osnovi koda MD5 NEDLIB Harvester utvrđuje je li određena stranica već bila zahvaćena. Nije sposoban preuzeti dinamične URL-e niti zaustaviti kružno preusmjeravanje poveznica (kružne klopke). Može prikupiti animacije u Flashu, iako nije sposoban prepoznati URL-e koji su ugrađeni u kod. To znači da ćemo, ako kliknemo na poveznicu u arhiviranoj kopiji, dobiti poruku da ne može naći odgovarajuću datoteku.

Da bi sačuvao prostor u računalnoj memoriji, arhiv je komprimiran u datoteci *.ZIP. Postupak dekomprimiranja je veoma brz. Neugodnost pri korištenju tim sustavom datoteka je u tome što traži sustav za skladištenje dodatne aplikacije za indeksiranje tekstova prikupljenih dokumenata.

Rješavanje tog problema bio je osnovni cilj projekta *Nordic Web Archive* (NWA), u okviru kojega su 2002. godine konačno razvili program za indeksiranje tekstova koji bi se koristio u kombinaciji s već postojećim aplikacijama za arhiviranje mrežnih stranica, među njima i s robotom NEDLIB Harvesterom. Dokumenti se čuvaju na trakama, a ne u bazi podataka jer bi to usporilo proces smještanja. Za indekse je potrebna posebna baza.

NEDLIB Harvester bio je najrasprostranjeniji robot za prikupljanje mrežne građe u Europi. Međutim, pokazivao je sljedeće pogreške u radu:

- nije bio sposoban prepoznati dugotrajne ispade u lokalnim mrežama - u tim ispadima proces prikupljanja bio je zamrznut i sustav je blokirao;
- nije bilo moguće podesiti različite dubine prikupljanja za različita mrežna mjesta;

- bio je "neugodan" sa poslužiteljima i pokazivao je veliku nestabilnost.

U Moravskoj knjižnici Brno prilagodili su NEDLIB Harvester svojim potrebama, a koristili su se njime za prikupljanje češke mrežne građe. Očekivali su da će skandinavske zemlje biti zainteresirane za njegov daljnji razvoj. Međutim, umjesto toga, članovi konzorcija IIPC započeli su s razvojem novoga alata za prikupljanje mreže, Heritrixa.

WGET je robot sa otvorenim kodom (*open source*) koji su razvili u Danskoj za projekt "netarkivet.dk".¹²⁸ Prikupljanje mrežne građe započinje iz početnog popisa URL-a i nastavlja proces preko postojećih poveznica na tim stranicama. Omogućava ograničavanje prikupljanja na različite načine, naprimjer može se odrediti koliko poveznica želimo slijediti iz početnog URL-a ili možemo ograničiti prikupljanje na određenog "hosta". Moguće je odrediti i vremenski razmak između dvaju preuzimanja. Vanjske poveznice može preuzeti tako da pokazuju na polaznu poveznicu (originalni dokument).

WGET svladava statične i dinamične URL-e jednakom brzinom.

WGET, kao ni NEDLIB, nije sposoban zaustaviti kružne klopke. Može preuzeti animacije u Flashu, iako ni on nije sposoban prepoznati URL-e koji su ugrađeni u kod. I tu se pojavljuje poruka da ne pronalazi datoteke.

HTTrack Web Site Copier

HTTrack je također robot s otvorenim kodom i dostupan je na mreži.¹²⁹ Napisan je u programskom jeziku C i prilagođen je za platforme Windows (95/98/NT/2000) i Unix. Verzija Windowsa (WinHTTrack) ima grafičko korisničko sučelje (GUI - *Graphic User Interface*).

Karakteristike HTTrack su sljedeće:

- snima mrežno mjesto za neumreženo (*offline*) pregledavanje;
- osuvremenjuje zrcalna mrežna mjesta;
- može produžiti i završiti prekinuto prikupljanje mreže;

¹²⁸ Dostupan je na URL: <ftp://ftp.gnu.org/gnu/wget/>. (15.12.2006.)

¹²⁹ HTTrack. URL: <http://www.httrack.com>. (15.12.2006.)

- ima veliki broj pravila (*scan rules*) za uključivanje i isključivanje URL-a, poveznica ili tipova datoteka;
- može se namjestiti dubina prikupljanja, veličina datoteka i veličina mrežnog mjesta;
- ima razne mogućnosti za namještenje strukture lokalnoga direktorija - osnovno podešavanje slijedi strukturu mrežnoga mjesta;
- može se namjestiti vrijeme početka prikupljanja;
- poštuje pravila *robots.txt*, a može ih i mimoići;
- može upotrijebiti pohranjene *cookies*;
- dostiže veliki broj poveznica u najvećoj brzini prikupljanja;
- ima potporu proxy (veze) s opcijom provjere autentičnosti (*authentication*);
- ima mogućnost upravljanja s pauzama i minimalni stupanj prijenosa;
- sposoban je automatski promijeniti poveznice URL-a originalnih mrežnih stranica u relativne poveznice u strukturi lokalnoga direktorija; prednost je to da se strukture direktorija zrcalnih mrežnih mjesta mogu preslikati u druge lokacije i da se mogu normalno pregledavati.¹³⁰

4.2 Softverske aplikacije za pristup građi

Kod projekata Kulturarw3 i NEDLIB mogle su se samo arhivirati mrežne stranice, a nije bio moguć korisnički pristup arhivu. Projekt *Nordic Web Archive* (NWA)^{131,132} riješio je taj problem. U suradnji s poduzećem FAST Search & Transfer ASA nordijske su nacionalne knjižnice razvile prototip sustava za pristup web arhivu. Taj je sustav neovisan o unutarnjoj strukturi i arhitekturi web arhiva te se može prilagoditi različitim okruženjima. U projektu su surađivale nacionalne knjižnice Danske, Finske, Islanda, Norveške i Švedske, koje su samostojno čak i sa različitim sustavima već zahvaćali mrežnu građu. Projekt je financirao NORDUnet2,¹³³ zajednički istraživački program nordijskih zemalja s ukupnim sredstvima u iznosu od 250.000 €. Projekt je

¹³⁰ Tehnički podatci o HTTPTracku su preuzeti iz: Verhoeven, I. H. Archiving web publications, Dec. 2002.

¹³¹ Bryggjeld, S. A. Access to web archives: the Nordic web archive access project. // 68th IFLA Council and General Conference: Libraries for Life: Democracy, Diversity, Delivery, August 18th - 24th 2002, Glasgow, Scotland: [proceedings]. URL: [URL:http://www.ifla.org/IV/ifla68/papers/090-163e.pdf](http://www.ifla.org/IV/ifla68/papers/090-163e.pdf) (17.4.2003.)

¹³² Nordic Web Archive. Helsinki : NORDINFO. URL: <http://nwa.nb.no/> (15.4.2003.).

¹³³ NORDUnet2: URL: <http://www.nordunet2.org> (15.12.2006.) je istraživački program koji financira nordijski Savjet ministara i nordijska vlada.

trajao od rujna 2000. do lipnja 2002. godine. Kao rezultat nastalo je sučelje s modulom za pristup NWA (*NWA toolset*) koje je bilo namijenjeno pretraživanju i navigaciji po arhiviranoj mrežnoj građi.

Verzija 1.0 NWA Toolset bila je sastavljena iz sljedeća tri modula:^{134,135}

- **modul za dobivanje/preuzimanje dokumenata (hvatač - *document retriever*)**, koji je bio sučelje u web arhiv. To je ulazna točka za predaju dokumenata i njihovih metapodataka u arhiv. Sučelje se mora dodatno prilagoditi informacijskoj tehnološkoj infrastrukturi knjižnice u kojoj se namješta;
- **Eksporter - modul za izvoz arhivskih dokumenata** - Taj modul preuzima dokumente preko hvatača i priprema ih za indeksiranje. Obično su to popisi URI-a koje Eksporter mora procesirati. Dokumenti se pretvaraju u format HTML, odakle se automatski preuzimaju metapodatci i smještaju se u dokumentni format NWA, koji je u XML-u i sadrži indekse za pretraživanje;
- **modul za dohvat (access)**, koji je dijelio zajedničko sučelje s pretraživačem i modulom za pridobivanje dokumenata, što je omogućavalo traženje, pretraživanje, listanje ili navigaciju po web arhivu.

NWA nudi sljedeće funkcije:

- omogućuje pristup web arhivu uz pomoć običnoga sustava za pretraživanje, što znači da nisu potrebni posebni dodatci ("*plugins*") za njegovo djelovanje;
- postiže jednaku brzinu kao i sustavi za pregledavanje na mreži;
- njegovo se sučelje ne razlikuje od internetskoga, koje omogućava traženje po metapodacima i po sadržajima;
- omogućava pregledavanje po cijelim dokumentima;
- omogućava navigaciju po arhivu po vremenu.

¹³⁴ Hakala, J. Archiving the web : European experiences // Program: electronic library and information systems, 38, 3(2004), str. 182.

¹³⁵ Hallgrímsson, Þ. ; Bang, S. Nordic web archive. // IWAW 2003. URL: <http://bibnum.bnf.fr/ecdl/2003/proceedings.php?f=hallgrimsson>. (15.12.2006.)

Osim spomenutih funkcija, FAST-ov pretraživač poznaje različite jezike i obavlja morfološku analizu riječi.¹³⁶ Modularnost omogućava lakše prilagođavanje različitim sustavima.

Svaka je knjižnica u projektu izgradila svoje sučelje. Prije završetka projekta 2002. godine, nordijske su nacionalne knjižnice imale, za svoju internu upotrebu, kompletan sustav za sakupljanje, arhiviranje i indeksiranje mrežne građe. Da bi aplikaciju NWA mogle iskoristiti i druge knjižnice, inicirali su nastavak projekta u NWA II, koji bi omogućio svim zainteresiranim knjižnicama upotrebu ovoga alata i omogućio njegov daljnji razvoj. U ožujku 2003. godine započeo je projekt NWA II, koji je također financirao NORDINFO, a završen je u ožujku 2004. godine. U tom su razdoblju alat prilagodili različitim programima za indeksiranje mrežne građe. Razvili su ga u sustav s otvorenim kodom (*Open Source SW*), dodali su mu funkcije koje su zahtijevali istraživači i drugi korisnici te ponovno s njima ocijenili sučelje. Alat je dostupan na URL: <http://nwatoolset.sourceforge.net/> od siječnja 2004. godine.

Projektna skupina NWA sudjeluje s *Internet Archive* u konzorciju IIPC tako da su već pripremili novo sučelje za pristup, WERA (WEb aRchive Access) (http://nwa.nb.no/wera/index.php?query=Norway&year_from=&year_to=), koje koriste s novim robotom za prikupljanje mreže Heritrix. Aplikacija za traženje dostupna je na URL: <http://nwa.nb.no/>.

4.3. Razvoj u okviru konzorcija IIPC (*International Internet Preservation Consortium*)

Heritrix

Krajem 2002. godine *The Internet Archive* osjetio je potrebu da zbog suradnje s raznim partnerima i svojih vlastitih potreba prikupi i arhivira svjetsku mrežu u novim oblicima. U tu su svrhu odlučili razviti nov alat za prikupljanje svjetske mreže. To bi bio robot napisan s otvorenim kodom (*open source*), što bi omogućilo suradnju s ostalim institucijama.

¹³⁶ Hakala, J. Archiving the web : European experiences // Program: electronic library and information systems. 38, 3(2004), str. 181.

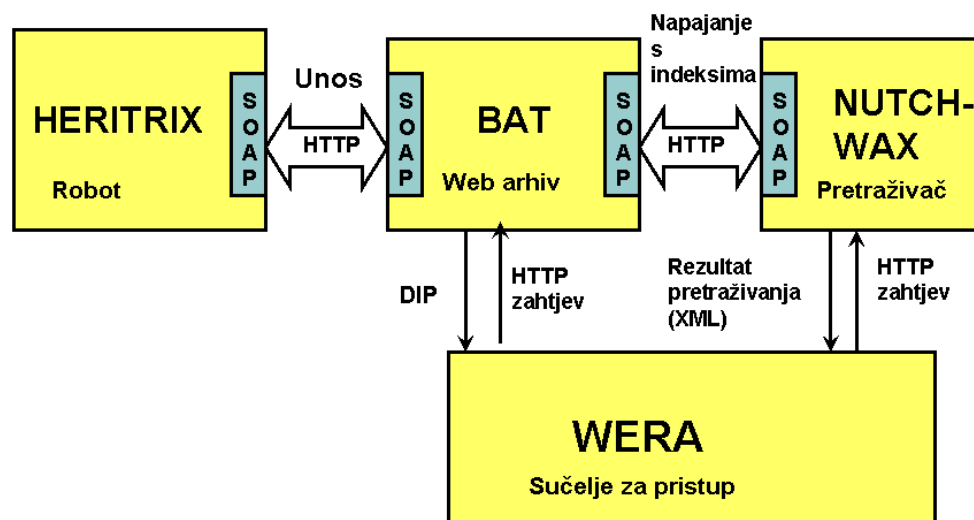
The Internet Archive započeo je s razvojem novog robota sakupljača početkom 2003. godine i nazvao ga Heritrix.¹³⁷ Alat je napisan u Javi, jer Java omogućava modularni dizajn i komponente koje se mogu raširiti i individualno nadomjestiti, a usto je to rasprostranjen jezik kvalitetnih knjižničnih aplikacija s otvorenim kodom. Heritrix bi trebao biti sposoban izvoditi različite tehnike prikupljanja mreže, kao što su: *broad crawling*, *focused crawling*, *continuous crawling* te *experimental crawling*. Robot je dostupan i ima besplatnu licencu, *Gnu Lesser General Public Licence*.

Europske su knjižnice 2002. godine započele pregovore s *The Internet Archive* o mogućnosti zajedničke suradnje u razvoju novoga robota za arhiviranje mreže. Budući da *Finnish Centre for Scientific Computing*, zbog nedostatka sredstava, nije više bio zainteresiran za daljnji razvoj i održavanje NEDLIB Harvester, i u finskoj nacionalnoj knjižnici su se odlučili na suradnju sa *The Internet Archive*. Na proljeće 2003. godine osnovali su konzorcij *The International Internet Preservation Consortium* (IIPC). Njegov je cilj bio razvoj standarda, dobre prakse i programa (robota) za arhiviranje mrežne građe. IIPC-u je, uz *The Internet Archive*, pristupilo još 12 nacionalnih knjižnica.¹³⁸ U srpnju 2003. godine u Parizu su potpisali osnivački ugovor za trogodišnje razdoblje. U svoje djelovanje planiraju uključiti i ostale nacionalne knjižnice. Osnovali su šest radnih skupina, koje su preuzele različite aspekte web arhiva. *The Internet Archive* je već u prvoj polovini 2003. godine započeo s razvojem novog alata za preuzimanje mreže, Heritrixa.¹³⁹ Među njihovim su sponzorima Hewlett i Fondacija Sloan, Etree, dobrotvorna ustanova koja besplatno snima glazbene koncerte uživo. Godine 2004. pozvali su na suradnju nordijske europske knjižnice, i to zbog njihovih bogatih iskustava na projektu NWA (*Nordic Web Archive*), u kojem su izradili sučelje i kompletni modul za pristup web arhivu, WERA. IIPC je iskoristio ta iskustva, kao i njihovu definiciju arhivskoga formata za smještanje metapodataka, WARC. Sustav, koji grade u okviru IIPC-a, prikazan je na **slici br. 2.**

¹³⁷ Heritrix potječe iz stare grčke riječi *heiress*, što znači "žena, koja nasljeđuje". Ime aludira na ulogu koju ima Internet Archive on nastoji sačuvati digitalnu kulturu za buduće generacije.

¹³⁸ U IIPC-u sudjeluju nacionalne knjižnice Australije, Kanade, Danske, Finske, Francuske, Islanda, Italije, Norveške, Švedske, Velike Britanije, Kongresna knjižnica SAD-a i The Internet Archive. Članstvo je tada bilo ograničeno samo na partnere, koji su već bili povezani u prethodnim projektima. Početkom 2007. godine IIPC je primio u članstvo i ostale knjižnice.

¹³⁹ The Internet Archive ponudio je na mreži aplikaciju Heritrix s otvorenim kodom (*open source*). Dostupan je na URL. <http://crawler.archive.org/> (15.12.2006.)



Slika br. 2: Sustav koji razvija IIPC za web arhiv. (Izvor: Netpreserve.org: IIPC Web Archiving Toolset (PPT prezentacija na IAWW 2005))

Heritrix je testiran na Linuxu, iako se može upotrijebiti i na platformi Windows, Macintosh OS i ostalim operacijskim sustavima.

Prva javno dostupna verzija Heritrixa (0.2.0) završena je u siječnju 2004. godine. U lipnju 2004. godine Heritrix je postao formalna aplikacija za prikupljanje mreže skupine projekta NWA. U kolovozu 2004. godine izašla je verzija 1.0.0, koja je bila prilagođena za fokusirano prikupljanje mrežne građe (*focused crawling*).¹⁴⁰ Verzijom 1.2.0 (u studenom 2004.) postigli su veću "ljubaznost" do poslužitelja (vodi brigu da svojim djelovanjem ne optereti "hostove" koje obilazi) s mogućnošću vremenskog namještanja intervala prikupljanja. Verzija 1.4.0 podupire prikupljanje mreže s rastućom strategijom (*incremental crawling*).¹⁴¹ Robot može u svakom prikupljanju preuzeti više od dvadeset milijuna mrežnih stranica i može raditi bez prekida tjedan ili više dana.

¹⁴⁰ Mohr, G. ...[et al.]: An introduction to Heritrix : an open source archival quality web crawler. // 4th International Web Archiving Workshop (IAWW 2004). URL: <http://www.iwaw.net/04/Mohr.pdf>. (15.12.2006.)

¹⁴¹ Stack, M. Heritrix crawler update. PPT presentation presented at 5th International Web Archiving Workshop (IAWW05), 22-23 September 2005, Vienna, Austria. URL: <http://www.iwaw.net/05/stack1.pdf>. (15.12.2006.)

Na međunarodnoj radionici IWAW 2006. godine predstavljena je nova verzija 1.10.0 Heritrixa, koja djeluje na Javi 1.5/5.0. Ova je verzija brža i ima veći kapacitet prikupljanja, bolju automatsku kontrolu sadržaja, mogućnost boljeg izlučivanja duplikata i nerelevantnih sadržaja te mogućnost višestupanjske koordinacije prikupljanja s većega broja računala. Smanjuje troškove pohrane nepromijenjenih sadržaja i podupire nove protokole i formate, kao što su pokretne slike (*Streaming Media*) te format WARC. Na mreži je u siječnju 2007. godine objavljena nova verzija 1.10.2 koja jednim prikupljanjem može prikupiti do dvije milijarde mrežnih stranica. Uz sve to, namjeravaju povećati broj članova konzorcija IIPC-a.

Kod razvoja Heritrixa programeri u *The Internet Archive* imali su za cilj postići prikupljanje uz pomoć rastuće strategije (*incremental crawling*).¹⁴² Modularnost omogućava da postizanje optimalnih rješenja kod svladavanja različitih tehnika i strategija prikupljanja mrežne građe. Proces prikupljanja mreže odvija se u sljedećih pet koraka:

1. izabire se URI u popisu planiranih mrežnih lokacija za prikupljanje
2. preuzima se URI
3. analiziraju se ili arhiviraju rezultati
4. odabire se zahvaćen URI koji je relevantan i dodaje se planiranim mrežnim lokacijama
5. provjerava se je li URI već zahvaćen te se ponavlja cijeli proces.

Najvažnije komponente Heritrixa su *Scope*, *Frontier* i lanac procesora koji imaju za cilj definiranje prikupljanja.

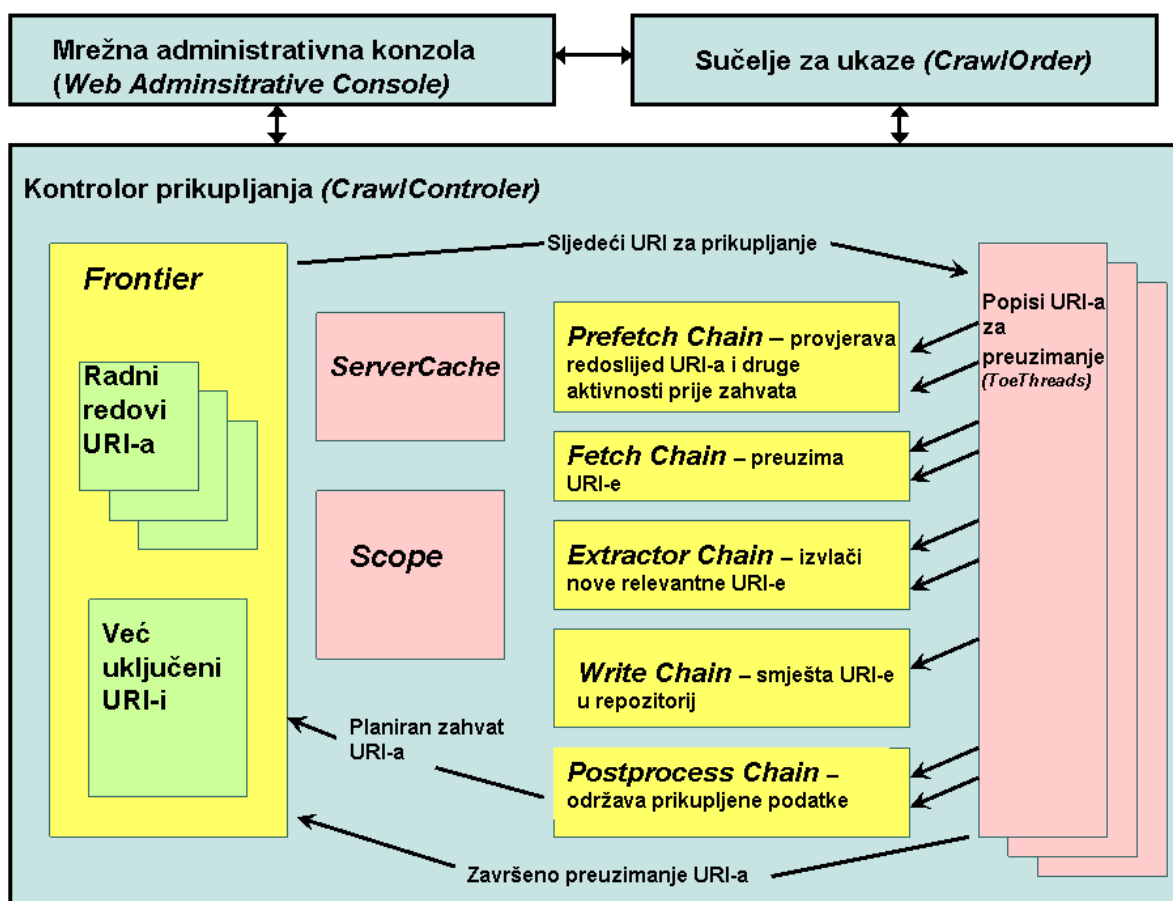
Scope određuje koji se URI-i slažu s kriterijima prikupljanja. Uključuje početne URI-e i pravila koja određuju relevantnost prikupljenih URI-a (4. korak).

Frontier slijedi popis planiranih URI-a i uspoređuje ih s onima koji su već zahvaćeni. Odlučuje o redoslijedu URI-a. (4. korak).

¹⁴² Sigurdsson, K. Incremental crawling with Heritrix. // 5th International Web Archiving Workshop (IWAW05), 22-23 September 2005, Vienna, Austria. URL: <http://www.iwaw.net/05/papers/iwaw05-sigurdsson.pdf>. (15.12.2006.)

Lanac procesora uključuje modularne procesore koji izvode specifične i lančane procese na svakom URI-u koji je na redu (koraci 2 - 4).

Arhitektura Heritrix prikazana je na slici br. 3.



Slika br. 3: Arhitektura Heritrix. (Izvor: Mohr, G., ... [et al.]: An introduction to Heritrix.... Figure 1.¹⁴³)

Heritrix je sastavljen od mrežnoga administrativnog sučelja (*Web Administrative Console*), koje je samostalna mrežna aplikacija na Jetty Java HTTP poslužitelju, zatim komandnoga sučelja (*CrawlOrder*), koji omogućava unos naredbi za preuzimanje, te od kontrolora prikupljanja mrežne građe (*CrawlController*), koji povezuje sve procese kod prikupljanja.

Komandno sučelje stvara *Scope*, koji zatim prosljeđuje *Frontieru* početne URI-e i potom mu poručuje odgovaraju li određeni URI-i postavljenim kriterijima odabira. Pritom vodi brigu da

¹⁴³ Mohr, G. ... [et al.] Op. cit. URL: <http://www.iwaw.net/04/Mohr.pdf> (15.12.2006.)

svojim djelovanjem ne optereti "hostove" koje obilazi. Ako određen URI odgovara kriterijima prikupljanja, onda on prelazi u lanac procesora koji ga preuzimaju, analiziraju i pohrane.

Programeri Heritrixa najprije su ugradili element *Frontier*, koji se po potrebi može lako uključiti kada treba održavati stanje prikupljanja. Drugi element je lanac procesora. Kad *Frontier* preda URI za prikupljanje, on prođe kroz seriju procesora. Po potrebi se mogu dodati novi procesori. Svaki procesor obavi određen zadatak, kao što je dobivanje dokumenta s interneta, izvlačenje poveznica ili upis njegova sadržaja na disk. Pri razvoju Heritrixa u *The Internet Archive* izradili su dva procesora: *ChangeEvaluator* i *Wait Evaluator*. Prvi utvrđuje je li se ponovno zahvaćen dokument promijenio te sprečava izvlačenje poveznica kod duplikata. Drugi procesor *WaitEvaluator* određuje interval između dvaju prikupljanja. On je ovisan o to je li se dokument u zadnjem prikupljanju mijenjao. U Heritrixi se može nastaviti taj interval za svakog "hosta" posebno. Vremensko namještanje vrijedi i kod formata datoteka koje preuzima, kao i za hijerarhiju stranica u mrežnom mjestu. Dodatno su razvili *Frontier* kao *AdaptiveRevisitFrontier* (*ARFrontier*), što omogućava stvaranje prioriternih lista u vrstama zahvaćenih URI-a za procesiranje. Kod običnog *Frontiera* vrijedi pravilo FIFO (*first in - first out*). Heritrix može istovremeno preuzimati po sto URI-a.

Da bi izbjegli prikupljanje bezvrijedne građe, dodatno su razvili procesor *HTTPContentDigest* i uključili ga u procesorski lanac između procesora za preuzimanje dokumenata i *ChangeEvaluator*a. Na taj se način automatski izdvajaju dokumenti koji unatoč promjenama nemaju arhivske vrijednosti. Dodatni filter *HTTPMidFetchUnhangedFilter* sprečava procesiranje duplikata u sustavu za prikupljanje.

Heritrix su testirali pri prikupljanja mrežnih stranica u nacionalnim knjižnicama Islanda, Francuske i Australije te u Nacionalnom arhivu SAD-a (NARA).

Zbog svoje sposobnosti prilagođavanja različitim tehnikama i strategijama prikupljanja mrežne građe Heritrix je široko prihvaćen u nacionalnim knjižnicama s različitom praksom preuzimanja mreže. Dostupan je na mrežnim stranicama (<http://netpreserve.org>).

Sučelje za pristup web arhivu IIPC - WERA (*Web Archive Collection Access Tool* ili *WAC viewer applicaton*)

Problem sučelja za pristup pohranjenoj građi u *The Internet Archive* (*Wayback machine*) bio je u tome da sa njim nije bilo moguće pretraživati po sadržaju arhiva. Zato su *The Internet Archive* i IIPC u razvoju sučelja za pristup mrežnim stranicama upotrijebili sustav s otvorenim kodom (*open source*) Nutch.¹⁴⁴ Njegove pozitivne strane su sljedeće:

- to je sustav s otvorenim kodom - aplikacija za pretraživanje na mreži, rezultat istraživačkoga projekta Apache;
- izgradili su ga na aplikaciji Lucene;
- može se upotrijebiti globalno, lokalno ili osobno;
- odabrali su ga za platformu pretraživača WAC (*Web Archiving Collections*);
- omogućava osnovne funkcije pretraživača, može se raširiti i dopuniti te nije skup, a može indeksirati više od 100 milijuna dokumenata.

Svi procesi obavljaju se po određenom vremenskom redoslijedu, a pretraživač Nutch može u slučaju vanjskih utjecaja (ispad sustava, nedjelovanje poslužitelja, slaba konfiguracija itd.) privremeno prekinuti pretraživanje. Nutch omogućava da se poslije prekida nastavi s procesima. Pretraživanje je slično kao i kod Googlea. Indeksiranje se izvodi tako što se indeksi i kontrolni zbrojevi URL, MD5, smjeste u posebnu bazu. Kod pretraživanja se uspoređuju MD5, pri čemu se odvajaju duplikati, a URL-i se sjedinjuju u skupine.

Može se reći da Nutch djeluje kao Google, jer omogućava pretraživanje s izlučivanjem duplikata. Heritrix smješta mrežne sadržaje u formatu ARC.

Bitan je podatak i promjena stranica u vremenskom redoslijedu. U tu svrhu Nutchov nadnevak u obliku YYYYMMDD zamijenili su s četrnaestoznamenastim brojem koji koristi *The Internet Archive Wayback Machine*: YYYYDDMMHHSS. Da bi omogućili pretraživanje po vremenskom redoslijedu, dodali su još i informaciju o lokaciji: zbirka, imeARC i ARCotisak. Duplikati se

¹⁴⁴ Stack, M. Full text search of web archive collections, IAWW 2005. URL: <http://archive-access.sourceforge.net/projects/nutch/iwaw/iwaw-wacsearch.pdf>

otpremaju u trenutku pretraživanja, a ne u trenutku indeksiranja. Funkcije koje Nutch omogućava su sljedeće:

- *sort* - sortiranje rezultata po željenim poljima
- *reverse* - otprema sortiranja
- *dedupField* - oznaka polja koja ne želimo duplicirati - stalna vrijednost (*default*) je mrežno mjesto
- *hitsPerDup* - informacija o količini dupliciranih polja
- *date* - četrnaestoznamenasti nadnevak
- *arcname* - ime datoteke ARC, u kojoj su rezultati pretraživanja
- *arcoffset* - početni otisak ARCname
- *collection* - zbirka kojoj pripadaju mrežna mjesta
- *exacturl* - pretraživanje po točnom URL-u.

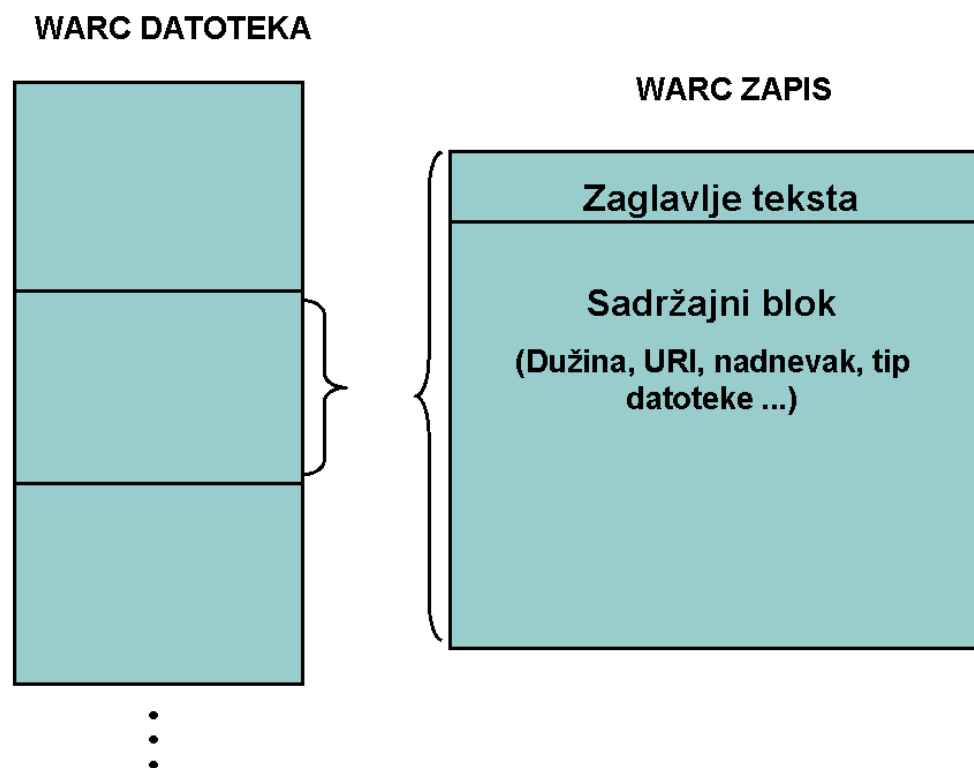
S proširenjima se Nutch može upotrebljavati za pretraživanje po zbirkama koje sadrže do 100 milijuna dokumenata.

U razvoju sučelja za pristup web arhivu WAC ima iste funkcije koje imaju Nutch od Lucene sa svim dodatcima (NutchWAX, odnosno Nutch, i Web Archive eXtensions) te NWA Access Tool. Novonastalo sučelje WERA omogućava pretraživanje po punom tekstu i prikazivanje mrežnih stranica po vremenskom redoslijedu. Verzija iz rujna 2005. godine pripremljena je za manje web arhive koje sadrže od 1 milijun do 10 milijuna dokumenata.¹⁴⁵

WARC (*Web ARChive file format*)

WARC je razvio IIPC na osnovi formata ARC, koji je izradio *The Internet Archive*. Od 1996. godine *The Internet Archive* napravio je više od 4 TB ARC-a. Datoteka WARC običan je redoslijed sadržajnih blokova, koji započinju s kratkim tekstovnim zaglavljem (*text header*) (vidi **sliku br. 4**)

¹⁴⁵ Masanès, J. ; Stack, M. *WEA : a web archive collection access tool*. PPT presentation at IAWW 2005, Vienna, 22 Sept. 2005. URL: <http://1.1.1.1/281420860/255225816T070216082734.txt.binXMysM0dapplication/pdfXsysM0dhttp://www.iwaw.net/05/stack2.pdf>. (15.12.2006.)



Slika br. 4: Anatomija datoteke WARC. (Izvor: Kunze, J. A. WARC: an archiving format for the web, 22.9.2006, IAWW 2005.)

Karakteristike formata WARC su sljedeće:¹⁴⁶

- može smjestiti bilo kakve metapodatke (predmetni klasifikator, jezik teksta, kodiranje itd) zajedno s drugim podacima;
- sadrži potporu za komprimiranje podataka i održavanje integriteta podatkovnoga zapisa;
- može smjestiti sve kontrolne informacije iz protokola za preuzimanje, naprimjer naslove naredbi (*request headers*), a ne samo odzive (*response information*);
- može smjestiti rezultate migracije podataka i povezati ih s ostalim smještenim podacima;
- može smjestiti podatke o duplikatima;
- može globalno smjestiti jedinstvene identifikatore zapisa.

¹⁴⁶ Kunze, J. WARC: an archiving format for the web. // IAWW 2005. URL: <http://1.1.1.1/285255580/253599752T070216082801.txt.binXsysM0dapplication/pdfXsysM0dhttp://www.iwaw.net/05/kunze.pdf>. (22.9.2006)

Postoji osam vrsta zapisa WARC:¹⁴⁷

- 1) **warcinfo** - obično je to prvi zapis u datoteci WARC, koji opisuje prikupljanje mreže (naprimjer dubinu, vremenska i druga namještenja, cilj, URL niša itd);
- 2) **odziv** (*response*) - odzivni zapis sadrži cijeli protokol odziva, kao što je cijeli HTTP odziv, uključujući zaglavlje (*headers*) i tijelo sadržaja mrežnoga prikupljanja; često sadrži adresu lokacije IP mrežne građe;
- 3) **izvor** (*resource*) - sadrži mrežnu građu bez informacija o protokolu odziva - naprimjer datoteka, koja je preuzeta iz lokalno dostupnog repozitorija;
- 4) **zapis o zahtjevu** (*request*) - sadrži informaciju o načinu na koji je prvi zapis bio zahtijevan (primjerice HTTP zahtjev);
- 5) **metapodatkovni zapis** (*metadata*) - napravljen je da bi opisao, objasnio ili pratio preuzetu mrežnu građu s podacima koji nisu bili zahvaćeni u drugim vrstama zapisa; može pokazivati i na druge zapise ili formate zapisa;
- 6) **zapis o ponovnom prikupljanju** (*revisit*) - sadrži podatke o ponovnom preuzimanju sadržaja, koji je već bio arhiviran i kod kojeg blok sadržaja već sadrži sažetak podataka koji se odnose na prethodni zapis; upotrebljava se samo onda ako je potrebno ponovno posjetiti prethodni zapis, jer sadržaj ovisi o njemu;
- 7) **zapis o pretvorbi** (*conversion*) - sadrži alternativnu verziju zapisa sadržaja koja je nastala kao rezultat arhivskoga procesa; promjene mogu biti točno opisane;
- 8) **zapis o nastavljaju** (*continuation*) - označava dokument koji je nastavak nekog drugog, a koji se logično povezuje s prethodnim zapisom (iz druge datoteke WARC).

Primjer metapodatkovnoga zapisa WARC:

```
warc/0.8 395 metadata http://www.archive.org/images/logo.jpg
20050708010101 text/xml
http://ark.cdlib.org/ark:/13030/xt12rk835gm/_s
Related-Record-ID: http://ark.cdlib.org/ark:/13030/
xt12rk835gm
<?xml version="1.0"?>
<harvestmetadata
xmlns="http://archive.org/harvest/0.8/">
<discovered-via>http://www.archive.org</discovered-via>
<download-time-ms>565</download-time-ms>
</harvestmetadata>
```

¹⁴⁷ Kunze, J. Idem.

Predložili su i da format WARC postane ISO standard i o njemu povelj javnu raspravu 2006. godine.¹⁴⁸

The Internet Archive je 2004. godine započeo s migracijom podataka u treću generaciju hardvera Petabox. Taj se sustav temelji na Linuxu i njegov dizajn omogućava RAID memoriju za oko 2.000\$ po TB ili 2.000.000\$ po PB. Prva instalacija novog dizajna izvedena je u Amsterdamu u novom Europskom arhivu, u kojem je i planiran smještaj zrcalnoga arhiva iz San Francisca.

4.4. Repozitorij web arhiva

Postoje tri vrste arhivskih sustava:¹⁴⁹

- **lokalni datotečni sustav** - sve su poveznice pretvorene u relativne i HTTP navigacija obavlja se u lokalnom sustavu datoteka; služi za preslikavanje pojedinačnih mrežnih mjesta odnosno, manju količinu mrežnih stranica; u tu je svrhu pogodan robot tipa HTTrack, kojim se koriste Australci; problem je u tome što traži poseban sustav za upravljanje dokumentima, njihovo sređivanje i preimenovanje svakog od njih¹⁵⁰ za različite formate jednakih URL-a; ne može se raširiti i prilagoditi ostalim sustavima;
- **mrežni poslužiteljski arhiv** - datoteke preslikava u neku vrstu kontejnera (naprimjer u format WARC kod Heritrix) zajedno s originalnim poveznicama između dokumenata; one se ne mijenjaju i moguća je navigacija kao i u živom okruženju interneta; pristup arhivu omogućen je pomoću mrežnoga poslužitelja; služi za veću količinu prikupljene mreže; pogodna je aplikacija za prikupljanje tipa Heritrix te poseban sustav za indeksiranje; vrlo je prilagodljiv, može se raširiti, iako je njegova implementacija teška; nezgodno je i to što nije moguć direktan pristup arhiviranim datotekama;
- **sustav koji nije web arhiv** – čine ga dokumenti koji su preuzeti iz originalnoga hipertekstualnoga okruženja, koji su reorganizirani po drugoj logici; ovakav bi arhiv

¹⁴⁸ Information and Documentation : The WARC File Format (working draft proposed to ISO TC46/SC4) URL: <http://www.niso.org/international/sc4/N595.pdf>. (15.12.2006.)

¹⁴⁹ Masanès, J. IIPC Web Archiving Toolset. Netpreserve.org // IAWW 2005. URL: <http://www.nla.gov.au/webarchiving/MasanésJulien.ppt>. (15.12.2006.)

¹⁵⁰ Konvencije imenovanja u datotečnom sustavu mogu biti različite od onih koje su dozvoljene na mreži. Tako primjerice, sustav može biti osjetljiv na velika ili mala slova, na oznake za izlaz, mogu postojati i zabranjeni znaci itd.

izgradili ako bi, naprimjer, pretvorili sve mrežne stranice u PDF format; služi za arhiviranje specifičnih ne-mrežnih zbirki (naprimjer *streaming media*¹⁵¹), odnosno za dokumente koji nisu dio mrežne građe; omogućava integraciju s tradicionalnim sustavima OPAC ili s ostalim organizacijskim lokalnim zbirkama; u tom se sustavu gubi hipertekstualna struktura.

4.5. Metode prikupljanja mrežnih stranica

Nacionalne knjižnice i ostale depozitarne institucije, koje su svjesne velike važnosti pohrane mrežne građe i ispunjavaju uvjete digitalnih arhiva, pokušaju naći primjeren pristup kako bi nacionalnu produkciju na mreži pohranile za budućnost.

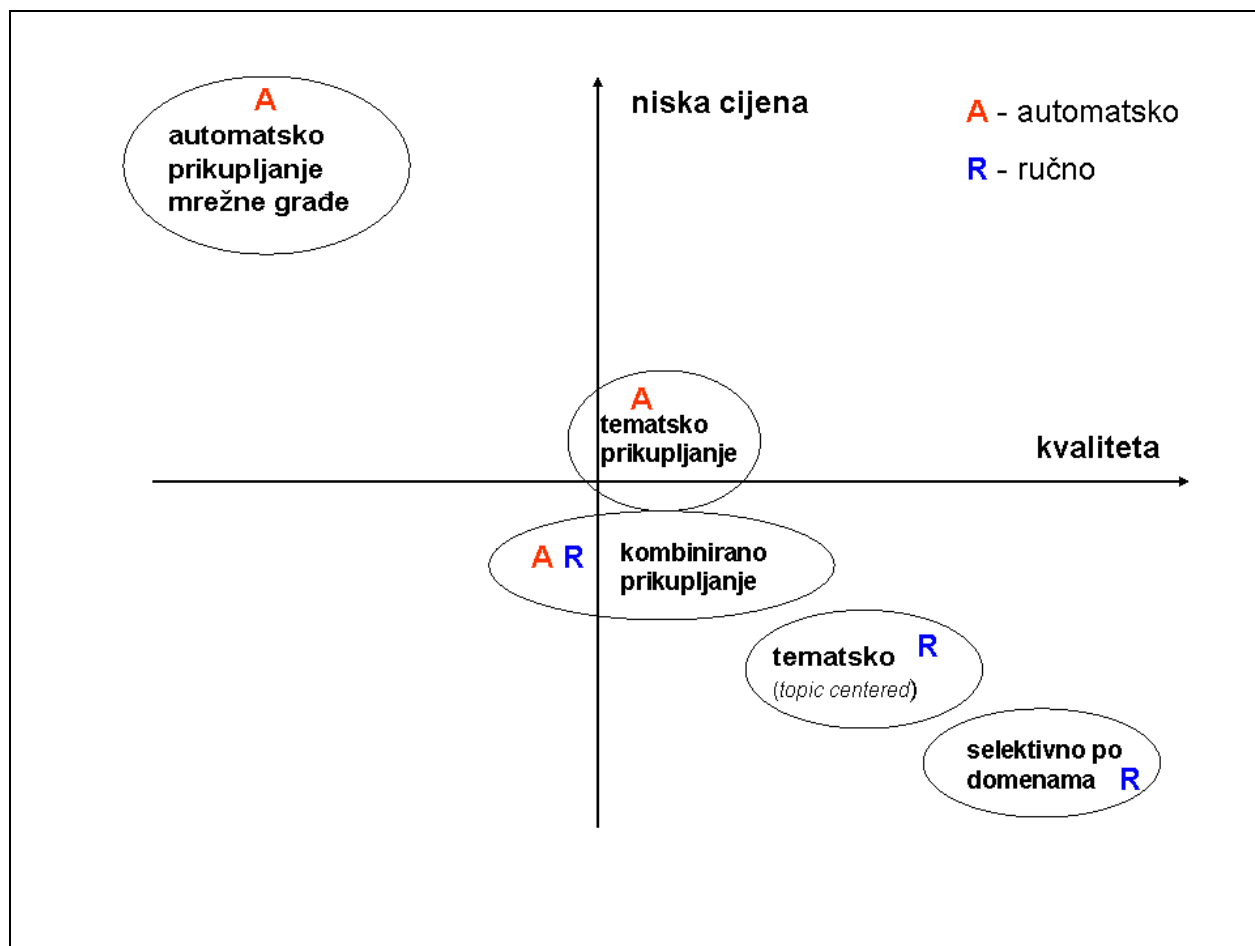
Julien Masanès 2005. godine u svojoj komparativnoj analizi¹⁵² utvrđuje da različiti pristupi arhiviranju mreže ovise o željenom obuhvatu, kvaliteti i metodama sakupljanja. (p.75). Što se tiče opsega, neke su se ustanove odlučile za prikupljanje mrežnih mjesta koja čine cjelovite jedinice građe (naprimjer australska Nacionalna knjižnica), ili za mrežnu građu koja je dio neke relevantne tematike ili događaja (naprimjer Kongresna knjižnica u SAD-u), dok ostale ustanove prikupljaju automatski na osnovi domene ili lokacije mrežnih stranica na određenim poslužiteljima, pri čemu je sadržaj podređen (naprimjer švedska Kraljevska knjižnica).

Projekti se razlikuju i s obzirom na metodološki pristup kod otkrivanja, pridobivanja i opisivanja sadržaja prikupljenih mrežnih mjesta, odnosno stranica. S jedne strane opažamo potpunu automatizaciju tog procesa, dok s druge strane imamo ručno prikupljanje i kolektivno odlučivanje o odabiru određenih dokumenata na mreži. Činjenica je da što se više prikuplja ručno, veća je i

¹⁵¹ Na radionici IWAW krajem rujna 2006. godine u Alicanteu, Španjolskoj, su predstavili nova dostignuća na području arhiviranje streaming media. To je donedavno bilo nemoguće. Vidi: Ashenfelder, M. Web harvesting and streaming media. // IWAW'06 : Proceedings of the 6th International Web Archiving Workshop, September 21-22 2006, Alicante, Spain / Masanès, J. ; Rauber, A., str. 125-145; Baly, N. ; Sauvin, F. Archiving streaming media on the web : proof of concept and first results. IWAW'06 : Proceedings of the 6th International Web Archiving Workshop, September 21-22 2006, Alicante, Spain / Masanès, J. ; Rauber, A., str.147-181. Referati su dostupni na URL: <http://www.iwaw.net/06/>. (15.12.2006.)

¹⁵² Masanès, J. *Web archiving methods and approaches: a comparative study*. // Library trends, 54, 1(2005), str. 72-90.

subjektivnost pri odabiru. Ručno prikupljanje mrežne građe zahtijeva veći broj stručnjaka, više sredstava i više vremena. Pri automatskom prikupljanju mrežne građe potrebna su dodatna sredstva samo za prostor i upravljanje mrežnim arhivom. Između tih dviju krajnosti nalazi se i tematsko prikupljanje (vidi **sliku br. 5**) koje može biti ručno ili automatsko.

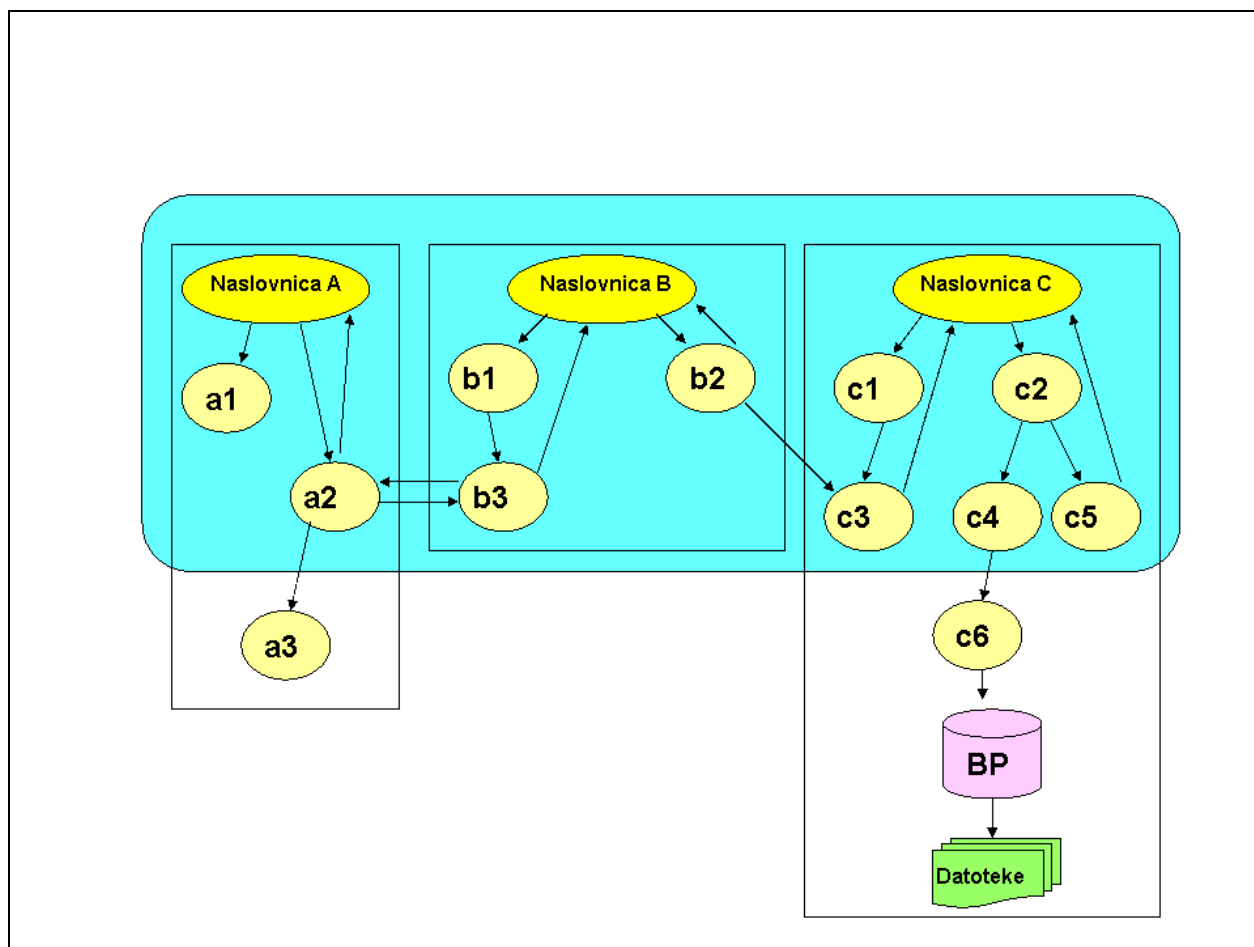


Slika br. 5: Odnos cijene i kvalitete pri automatskom i ručnom prikupljanju mrežne građe. (Izvor: Masanès, J. *Web archiving methods and approaches: a comparative study*. // *Library trends*. 54, 1(2005), str.72-90)

Kvaliteta, po Masanèsu, može ovisiti o aktivnim poveznicama u preuzetim dokumentima, o mogućnostima pristupanja te o sadržaju mrežnoga arhiva. Što je automatiziranije prikupljanje, teže se osigurava veća kvaliteta. Zato su stranice koje su ručno preuzete mnogo kvalitetnije, cjelovitije i imaju više mogućnosti da kao takve prežive i u budućnosti.

Pri automatskom prikupljanju mreže roboti započinju preuzimanje građe preko popisa ulaznih domena i preko njihovih poveznica šire se na ostale stranice i mrežna mjesta dok se proces

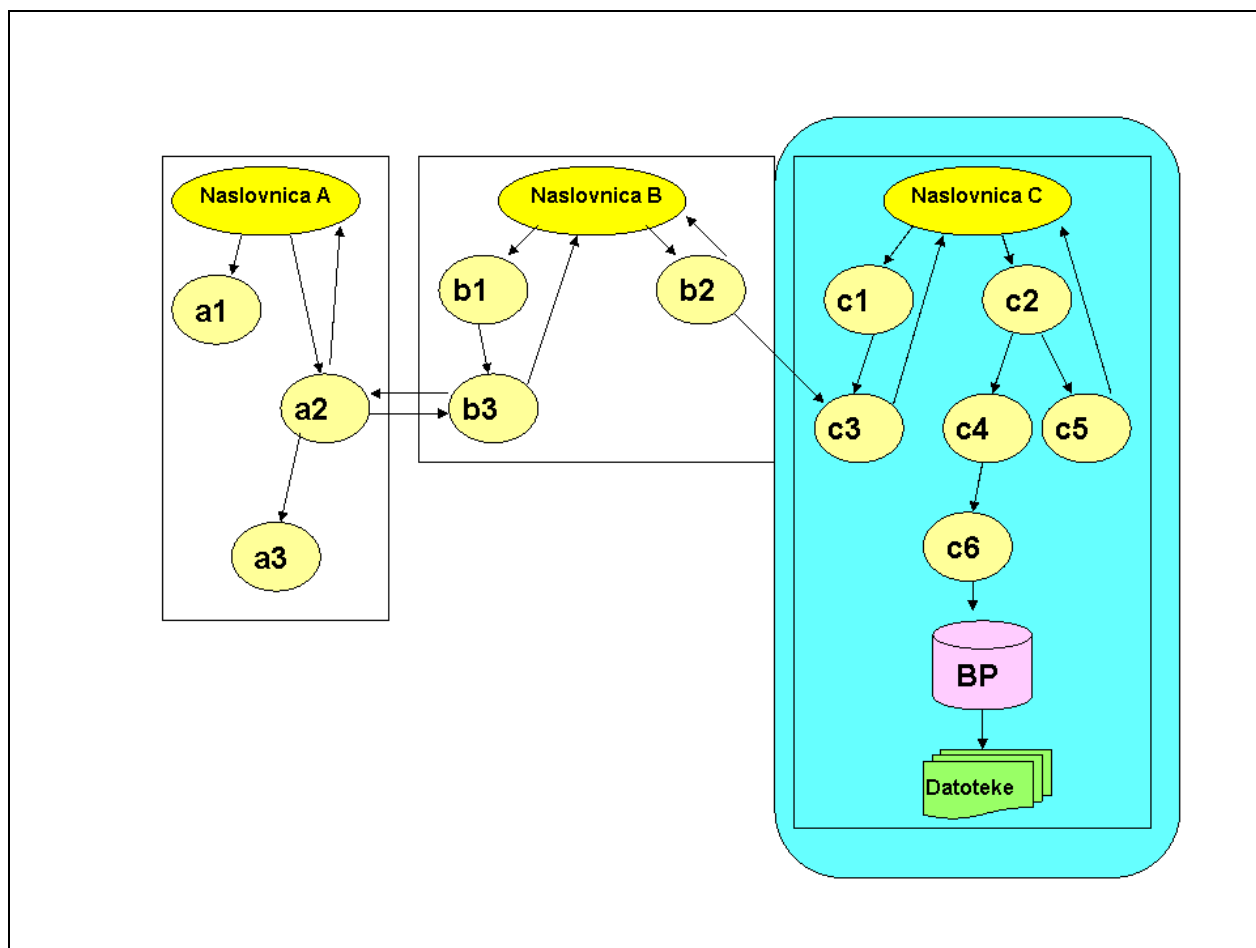
prikupljanja ne završi. Kod tih robota možemo nastaviti dubinu preuzimanja, recimo do treće ili četvrte razine, ili broj stranica po mrežnom mjestu. Takvo prikupljanje nazivamo "horizontalno" ili "ekstenzivno". Problem kod ovog tipa sakupljanja je taj da se ne preuzima sve što se nalazi ispod određenih razina. Zbog toga, kod pregledavanja mrežnih stranica koje imaju poveznice na niže razine, one neće djelovati ili će im nedostajati određeni elementi, kao što su slike ili logotipi. (vidi sliku br. 6)



Slika br. 6: Primjer ekstenzivnoga prikupljanja mrežne građe, kod kojeg nisu bile uključene mrežne stranice a3 i c6. (Izvor: Masanès, J. Web archiving methods and approaches: a comparative study. // Library trends, 54, 1(2005), str.72-90)

Ako pak odaberemo određene URL-e i namjestimo robot ili pauk da preuzima samo one mrežne stranice koje su na određenom poslužitelju, onda imamo primjer "vertikalnoga" ili "intenzivnoga" prikupljanja. (vidi sliku br. 7). U ovom slučaju može se dogoditi da robot previše opterećuje poslužitelje i prouzrokuje teškoće u njihovu djelovanju. Intenzivni način prikupljanja mrežne

građe karakterističan je za prikupljanje po mrežnim mjestima, odnosno za tematsko prikupljanje. Ulazne su točke naslovnice mrežnih mjesta.



Slika br. 7: Primjer intenzivnoga preuzimanja mrežne građe. (Izvor: Masanès, J. *Web archiving methods and approaches: a comparative study*. // *Library trends*. 54, 1(2005), str.72-90)

Baeza-Yates i Castillo¹⁵³ su tijekom istraživanja modela vjerojatnosti korisničkoga pretraživanja utvrdili da pri preuzimanju mrežnih stranica do treće ili četvrte razine uspijevamo prikupiti oko 90% svih stranica koje obično posjete korisnici, odnosno stranica koje imaju bilo kakvu korisničku vrijednost. Istraživanje je temeljilo na pretpostavci da korisnici ulaze u mrežno mjesto preko njihovih naslovnica, iako mnogi korisnici slijede poveznice na mnogo nižim razinama drugih mrežnih mjesta. Zato Masanès predlaže da se kombiniraju oba pristupa tako da se

¹⁵³ Baeza-Yate, R. ; Castillo, C. *Crawling the infinite web*, 2004. URL: http://1.1.1.1/272351300/301076704T060817125403.txt.binXMysM0dapapplication/pdfXsysM0dhttp://www.dcc.uchil.e.cl/~ccastill/papers/baeza_05_crawling_infinite_web.pdf. (15.12.2006.), str. 10.

ekstenzivnim preuzimanjem prikupi mrežna građa do pete razine dvaput godišnje, a da se u međuvremenu više puta intenzivno prikupe stranice koje su od većeg značenja za depozitarnu ustanovu.¹⁵⁴ Ručno prikupljanje može povećati kvalitetu arhiviranih mrežnih mjesta koje su dio dubinske mreže, dok automatsko prikupljanje omogućava preuzimanje i onih stranica koje imaju veoma kratak životni ciklus, a iz sadašnje perspektive nemaju posebne vrijednosti, iako su dio društvenoga kulturnoga konteksta.

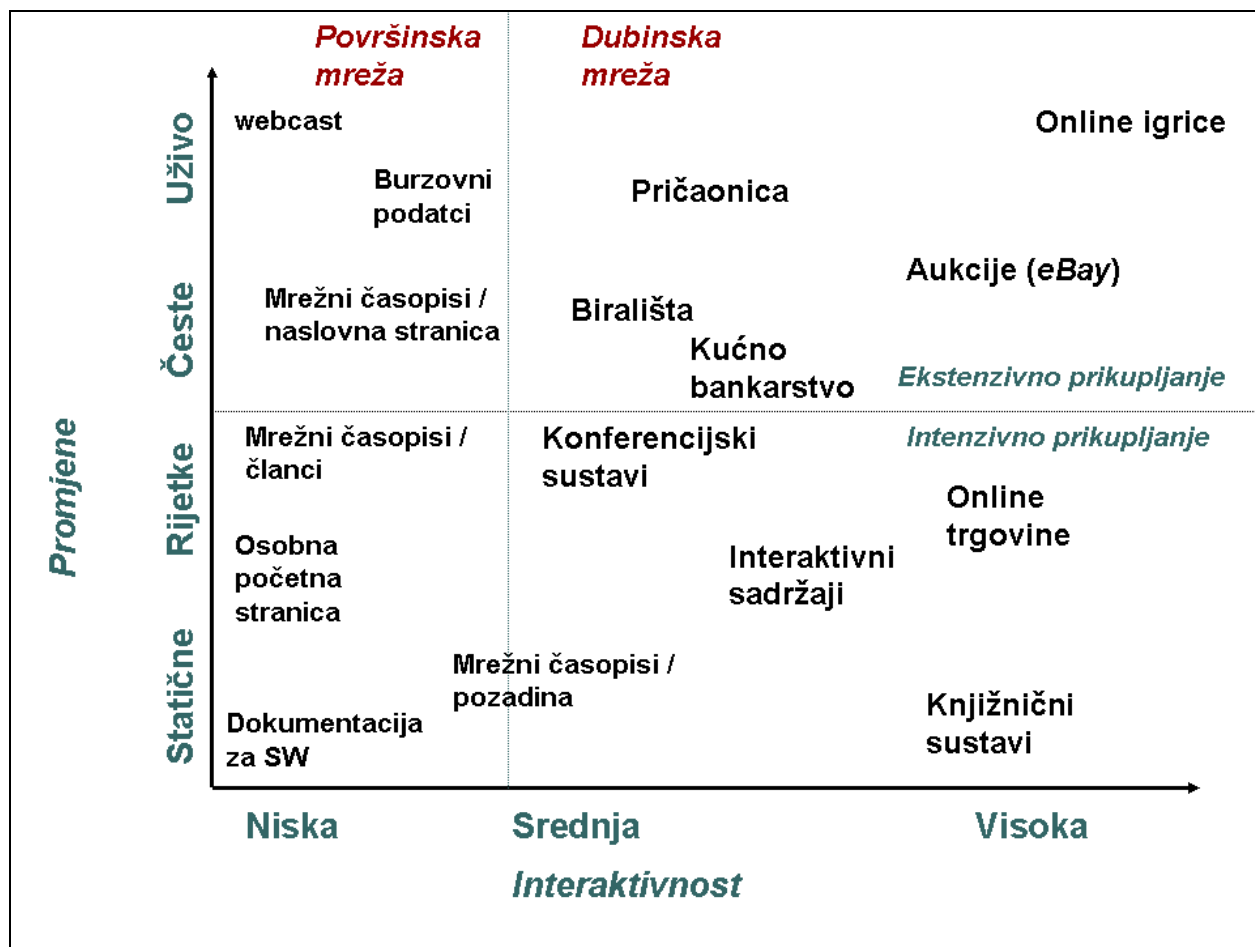
U svom je istraživanju Masanès uspoređivao dvije metode prikupljanja, automatsku, odnosno domensku (ekstenzivnu), i tematsku. Nacionalna knjižnica Francuske je 2002. godine pomoću robota HTTrack ručno prikupila mrežne stranice koje su bile sadržajno vezane uz izbore. Podatke su prikupljali od ožujka do svibnja 2002. godine. Iz cijele zbirke odabrali su 696 mrežnih mjesta vezanih uz predsjedničke izbore i 1002 mrežnih mjesta vezanih uz parlamentarne izbore. Uspoređivali su ih zatim s mrežnim mjestima koje su ekstenzivno prikupili u *The Internet Archive* (IA). IA je pokazao vrlo dobru pokrivenost mrežnih mjesta, iako su rezultati bili veoma slabi kod prikupljanja njihovih sekcija. Naime, kod ekstenzivnoga prikupljanja vrijeme odmicanja kod preuzimanje mrežne građe između dviju razina može onemogućiti cjelovito preuzimanje svih stranica jednoga mrežnoga mjesta. Dok robot preuzima stranice s više mrežnih mjesta na istoj razini, stranice na drugim razinama mogu se za to vrijeme mijenjati. Tako nastaje neusklađenost između već preuzetih mrežnih stranica i novih stranica na koje pokazuju njihove poveznice.

Odabir strategije prikupljanja uvelike ovisi o financijskim i tehničkim mogućnostima ustanove. Velik utjecaj ima i učestalost osuvremenjivanja relevantnih mrežnih mjesta, kao i složenost njihove interaktivnosti.

U izvještaju projekta "netarkivet.dk"¹⁵⁵ prikazana je matrična tabela različitih pristupa u odnosu na vrstu publikacija na mreži. (vidi **sliku br. 8**)

¹⁵⁴ Masanès, J. *Web archiving methods and approaches: a comparative study*. // Library trends, 54, 1(2005), str. 78.

¹⁵⁵ Idem.



Slika br. 8: Implementacija različitih pristupa pri preuzimanju mrežne građe u funkciji učestalosti mijenjanja i interaktivnosti mrežnih stranica. (Izvor: Izvještaj projekta "netarkivet.dk")

Na **slici br. 8** vidimo da na jednom mrežnom mjestu može biti različita dinamičnost promjena. Naprimjer kod mrežnih časopisa naslovnica je veoma promjenljiva, dok se njihovi članci rijetko mijenjaju.

Kod odabira metode prikupljanja mrežne građe treba posvetiti pažnju dvama bitnim aspektima: s jedne strane sadržaju mrežne građe, a sa druge alatima i infrastrukturi koji nam omogućavaju arhiviranje.

Prije odlučivanja o strategijama i metodama prikupljanja građe vrlo je važno odrediti što želimo prikupiti i sačuvati u web arhivu te na koji način želimo upotrijebiti taj sadržaj. U tom je smislu korisno detaljno odrediti kriterije odabira, kako u sadržajnom tako i u formalnom obliku.

S druge strane, prikazali smo kako djeluju različiti roboti, koja su svojstva elektroničke građe, kako je smještamo i kako joj pristupamo u web arhivima. Nakon određivanja kriterija odabira potrebno je obratiti pažnju na funkcionalnost robota prikupljača i način njegova djelovanja. Danas je već poznato da Heritrix postiže bolje rezultate od svih dosadašnjih alata za prikupljanje građe. Omogućava i ekstenzivno i intenzivno prikupljanje te radi na principu rastuće strategije. Time nam olakšava odabir metode i pomaže nam da se usredotočimo na sadržaj, a ne na formu. Naprimjer, ako želimo u cjelini prikupiti mrežno mjesto koje se češće mijenja, možemo namjestiti robot tako da prikuplja intenzivno i s većom učestalošću. Ako pak želimo dobiti što više primjeraka mreže u danom trenutku, koristit ćemo se ekstenzivnom metodom. U međuvremenu će se pojaviti teme ili događaji koji bi mogli biti zanimljivi kao cjelina i tom ćemo slučaju primijeniti tematski pristup. Kod intenzivnog i tematskog pristupa mogli bismo kombinirati ručno i automatsko prikupljanje. U svakom slučaju, odabir metode ovisi o tome što želimo sačuvati, odnosno što će biti važno iz perspektive budućega korisnika.

Na odabir metode prikupljanja može utjecati i broj stručnjaka koji u tome sudjeluju, odnosno njihova angažiranost u tom procesu.

U sljedećim poglavljima prikazane su vrste mrežne građe koje možemo naći na mreži i kriteriji njihova odabira.

5. TIPOLOGIJA MREŽNIH PUBLIKACIJA

Jedna od važnijih karakteristika elektroničkih publikacija je njihova višedimenzionalnost. U digitalnom mediju, konkretno na svjetskoj mreži, korisnik neprestano skače iz jedne dimenzije u drugu i preko različitih poveznica odabire relevantne sadržaje.^{156,157} Sadržaj i informacijska vrijednost postaju važniji od njihove forme. U bibliotekarstvu, koje operira tradicionalno priznatim konceptima, novi oblici sadržaja uvode zbrku. U želji da ih savlada, pokušava ih ukalupiti u već postojeće klasične oblike građe. Međutim, nova građa sadrži nove funkcionalnosti i premda su sadržaji slični, njihove funkcije ili vrijednosti mogu biti sasvim različite.

Osim toga, informacijsko-tehnološki razvoj omogućava nastanak novih nositelja, koji podupiru sve veće količine podataka, kao i novih računalnih programa i operacijskih sustava, koji lakše i brže upravljaju multimedijским sadržajima. Kao posljedica toga nastaju novi oblici publikacija, hibridni tiskani oblici, od kojih su neki već dobili ili tek trebaju dobiti svoj identitet. Tako naprimjer, dok nam je u okviru tradicionalne građe jasna definicija enciklopedije, u digitalnom je svijetu ponekad teško razlikovati multimedijšku enciklopediju sa zvučnim snimcima, obrazovnim igricama i posebnim softverom za pretraživanje od običnih obrazovnih igrica ili zvučnih snimaka popraćenih filmskim spotovima. Pojavljuju se i bibliografije, koje su sa svojim aktivnim poveznicama do punih tekstova prije popis članaka nego li popis bibliografskih referencija.

Jedno je od glavnih pitanja je li moguće iz perspektive naše generacije definirati nove tipove publikacija koje su nastale ili tek nastaju na mreži. Još uvijek razmišljamo u okviru tradicionalnih pojmova građe i teško je udaljiti se i pokušati iznova odrediti nove oblike, koje nalazimo na mreži. Mnogi su od tih mješavina već poznatih, drugi su sasvim novi. Tim problemom se danas

¹⁵⁶ Koncept dvodimenzionalnosti na mreži prvi put uvode Askehave, Nielsen i Ellerup u svome članku: What are the characteristics of digital genres? - Genre theory from a multi-modal perspective, p. 98a. In: Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS'05) - Track 4, 2005. <http://1.1.1.1/271155764/259797728T060814111756.txt.binXMysM0dapplication/pdfXsysM0dhttp://csdl2.computer.org/comp/proceedings/hicss/2005/2268/04/22680098a.pdf>

¹⁵⁷ Ridi navodi da "u analognom svijetu je dovoljan pogled za razumijevanje dokumenta kojim se bavimo; u digitalnom svijetu taj isti objekt može sadržati vrlo heterogene entitete i može biti više i češće hibridiziran." Više o tome u članku: The world as will and documentation. Definition, selection and access to remote electronic resources (RER). // Le risorse elettroniche : definizione, selezione e catalogazione = Electronic resources. Definition, selection and cataloguing. Proceedings of the International Conference, Rome, November 26-28, 2001 / uredio Guerrini, M. Milano : Bibliografica, 2002. Str. 439-451.

bave različite discipline, kao što su komunikologija, sociologija ili genologija interneta. Dok prve dvije pokušavaju razumjeti način komuniciranja i sociološki analizirati novonastale oblike izražavanja na internetu, genologija interneta pokušava prenijeti znanost o žanrovima na internet. Osim njih postoje i istraživanja o karakterizaciji interneta, koja također pokušavaju definirati i statistički odrediti sve oblike publikacija i komuniciranja na internetu.

Internet je istovremeno komunikacijski i publikacijski kanal. Mi se u ovom radu bavimo publikacijama. Pod publikacijom podrazumijevamo sve što je objavljeno, publicirano. U slučaju interneta govorimo o djelima koja su javnosti dostupna na mreži.¹⁵⁸ U australskoj Nacionalnoj knjižnici¹⁵⁹ navode slijedeću definiciju: "Publikacija je informacija, bez obzira na njezin format ili način dostave, tj. dostupna općoj ili određenoj publici, besplatno ili sa članarinom."¹⁶⁰

Osim tradicionalnih oblika i formi publikacija (elektroničke serijske publikacije, elektroničke knjige, digitalne slike itd.) na mreži nalazimo i ostale oblike, kao što su naslovnice poduzeća, interaktivne baze podataka, mrežni časopisi, blogovi itd. Stančić¹⁶¹ nabraja ove vrste građe, koja je dio elektroničke baštine: integrirajuće ili potpune elektroničke publikacije na bilo kom mediju, pretpublikacije (*preprints*), poslovni zapisi, baze podataka, elektronički nastavni materijali, softver, jedinstveni neobjavljeni materijali (materijali u elektroničkom obliku dostupni užem krugu korisnika - naprimjer pričaonice, forumi itd.), elektronički "rukopisi", zabavni proizvodi, umjetnička djela, dokumentarni zapisi i elektroničke reprodukcije.

¹⁵⁸ U ovom radu upotrebljavamo termin "publikacija" u značenju osnovne građe fondova knjižnica, za razliku od "dokumenta", koji su predmet arhivistike. Međutim, opredjeljenje publikacije dokumentom, s obzirom na to da je predmet dokumentalističke prakse, nalazimo kod mnogih autora. Primjerice Buckland, M. K. What is a "Document". // Journal of the American Society for Information Science 48, 9(1997), str. 804-809; Ridi, R. The world as will and documentation. Definition, selection and access to remote electronic resources (RER). // Le risorse elettroniche : definizione, selezione e catalogazione = Electronic resources. Definition, selection and cataloguing. Proceedings of the International Conference, Rome, November 26-28, 2001 / uredio Guerrini, M. Milano : Bibliografica, 2002. Str. 439-451.

¹⁵⁹ Online Australian Publications: Selection Guidelines for Archiving and Preservation by the National Library of Australia. Revised August 2005. URL: <http://pandora.nla.gov.au/selectionguidelines1003.html>. (15.12.2006.)

¹⁶⁰ "A publication is information, regardless of its format or method of delivery, that is made available to the general public, or to an identified public, either free of charge or for a fee." URL: <http://pandora.nla.gov.au/selectionguidelines1003.html> (15.12.2006.)

¹⁶¹ Stančić, H. Očuvanje elektroničkih informacijskih objekata : arhivi, knjižnice, muzeji - zajednička koncepcija" // 7. Seminar arhivi, knjižnice, muzeji... str. 26-35.

Postoji više podjela mrežne građe, koje se često isprepliću. Autori svoje podjele temelje na jednom kriteriju ili na više kriterija koji mogu biti formalni ili sadržajni:

Formalni kriteriji podjele mrežne građe

- **materijalna struktura građe** (mrežna stranica, mrežno mjesto, baze podataka na mreži, građa s programskom opremom i slično)
- **promjenljivost** (statične i dinamične publikacije itd.)
- **način nastanka** (građa koja je proizvod digitalizacije ili je nastala u digitalnom obliku)
- **funkcija građe** (osobne ili organizacijske naslovnice; katalozi, informacijski portali; baze podataka: telefonski imenici, bibliografske baze; transakcijske stranice: mrežna trgovina, bankovno poslovanje; računalne aplikacije itd.)
- **način pristupa** (publikacije sa slobodnim ili ograničenim pristupom, publikacije kod kojih je pristup omogućen preko indeksa, baze podataka itd.)
- **odgovorno tijelo ili stvaratelj** (mrežne stranice ili mjesta koje produciraju pojedinci, organizacije, društva, zakladnici itd.)
- **bibliografska obrada** (neomeđena i integrirajuća građa, knjige, serijske publikacije itd.).

Sadržajni kriteriji podjele mrežne građe

- **sadržaj građe** (beletristika, znanstveni tekstovi, mrežna trgovina, informacijski portali)
- **usredotočenost** građe (informativne, istraživačke, osobne stranice itd.)
- **pripadnost žanru** (naslovnice, blogovi, mrežne enciklopedije, RFC itd.)
- **informacijska funkcija** (dućanski prozori, mjesta za traženje i pretraživanje, mjesta s izvorima informacija itd.).

5.1. Formalna podjela mrežne građe

Najčešća formalna podjela mrežne građe na osnovi strukture građe na mreži je na mrežnu stranicu i mrežno mjesto.¹⁶²

Mrežna stranica je osnovna informacijska jedinica građe na mreži, koju čini jedna HTTP datoteka i obično se identificira jedim URL-om. Standard ISBD definira je kao stranicu "nekog hipertekstovnog dokumenta na World Wide Webu".^{163,164} Sličnu definiciju nalazimo i kod ostalih autora.¹⁶⁵

Za razliku od mrežne stranice, **mrežno mjesto** sastavljeno je iz skupa mrežnih stranica. Prime-Claverie, Beigbeder i Lafouge definiraju mrežno mjesto kao skupinu stranica sa zajedničkim ciljem i temom koje izrađuje i održava jedna te ista osoba ili organizacija.¹⁶⁶

Najpoznatija mrežna građa je **naslovnica** (*home page*). Ovdje se najčešće miješaju pojmovi mrežne stranice i mrežnog mjesta, jer naslovnica je samo ulazna mrežna stranica u mrežno mjesto ili u skup mrežnih stranica, a ne cijelo mrežno mjesto. Obično sadrži opće informacije o mrežnom mjestu ili ima funkciju kazala. Ipak, moguće je naći jednostavne naslovnice koje su jedna mrežna stranica.

U konzorciju OCLC¹⁶⁷ su u okviru projekta *Web Characterization*¹⁶⁸ ustanovili da mrežno mjesto možemo definirati na osnovi dvaju aspekata: materijalni aspekt, koji se odnosi na lokaciju na poslužitelju, i informacijski odnosno sadržajni aspekt, koji se odnosi na mrežno mjesto kao cjelovitu informacijsku jedinicu. Često se obje definicije preklapaju, iako postoje i primjeri kada se isključuju. Materijalna je definicija općenitija od informacijske, što se pokazalo kad su 2002.

¹⁶² O'Neill, E. T. ; Lavoie, B. F. Bibliographic control for the web. // The Serials Librarian. 37, 3(2000), str. 53-69.

¹⁶³ ISBD(CR), hrvatski prijevod str. 17.

¹⁶⁴ *Web page: One of the pages of a hypertext document in a World Wide Web site. Web pages, including the subset "home pages", refer to the huge collection of documents that make up the World Wide Web. (p.9)*

¹⁶⁵ Prime-Claverie, C. ; Beigbeder, M. ; Lafouge, T. Transposition of the cocitation method with a view to classifying web pages. // Journal of the American Society for Information Science and Technology. 55, 14(2004), str.1282-1289.

¹⁶⁶ Prime-Claverie, C. ; Beigbeder, M. ; Lafouge, T. Op. cit.

¹⁶⁷ Web sites : concepts, issues, and definitions. Dublin, OH : OCLC, January 1999. URL: <http://www.oclc.org/research/projects/archive/wcp/pubs/rn1-websites.htm>. (15.12.2006.)

¹⁶⁸ Web Characterization Project, URL: <http://www.oclc.org/research/projects/archive/wcp/default.htm> (15.12.2006.)

godine, u okviru spomenutog projekta, pomoću materijalne definicije utvrdili da je svjetska mreža sastavljena od 2,035,000 mrežnih mjesta, dok su na osnovi informacijske definicije pronašli oko 20 milijuna područja koja čine samostalne sadržajne jedinice. Dokazali su da je moguće da na jednom poslužitelju postoji više mrežnih mjesta, što roboti prikupljači ne mogu prepoznati.¹⁶⁹ Zato je mnogo prikladnija informacijska definicija, koja glasi: "mrežno mjesto je skupina sadržajno povezanih mrežnih stranica koje tvore sastavljeni objekt informacijske relevantnosti".¹⁷⁰

S obzirom na promjenljivost, mrežne stranice ili publikacije mogu biti statične ili interaktivne.¹⁷¹ Statična mrežna stranica je dokument koji je moguće čitati od početka do kraja bez potrebe da napušta stranicu i informacija je jednaka svim korisnicima. U većini slučajeva statične mrežne stranice sačinjavaju datoteke s produžetkom *.htm ili html. Interaktivna mrežna stranica je prilagođen dokument koji koristi vanjske programe da bi izvodio određene funkcije. Interaktivne mrežne stranice¹⁷² omogućavaju korisnicima ispunjavanje obrazaca, pretraživanje po bazama, listanje rezultata pretraživanja u određenom formatu i pristup zaštićenim mrežnim stranicama. Steenbakkera u izvještaju o projektu NEDLIB¹⁷³ interaktivne stranice naziva dinamičnim publikacijama jer se stalno mijenjaju ili su rezultat nekog informacijskog pretraživanja te ih roboti ne mogu u potpunosti prikupiti, već se dobiva samo njihov trenutni prikaz na mreži.

¹⁶⁹ U spomenutom dokumentu Web sites : concepts, issues, and definitions obrađuju se različiti primjeri mrežnih mjesta u kojima se uzima u obzir informacijska i materijalna definicija. Najjednostavniji su standardni primjeri, kod kojih je skupina mrežnih stranica dostupna iz jedne adrese IP. Zatim dolaze varijacije standardnog primjera, kao slučaj kad više adresa IP odgovara jednom domenskom imenu ili više domenskih imena odgovara jednoj te istoj adresi IP. Kod nekih primjera čak su našli različita domenska imena za jedno mrežno mjesto na istoj adresi IP. Mogući su i primjeri kad na istoj adresi IP postoje više mrežnih mjesta i/ili usluga koje imaju različita domenska imena (*virtual hosting*). OCLC nalazi također "logička mrežna mjesta", koja su u sastavu većih, složenih mrežnih mjesta. S druge strane mogući su primjeri kada jedno mrežno mjestu udružuje više adresa IP ili mrežu "*hostova*" u logičku cjelinu. Mrežna mjesta mogu biti postavljena na istoj adresi IP, iako su dostupna iz različitih ulaza (*port*).

¹⁷⁰ *Web Site (Information Definition): a set of related Web pages that, in the aggregate, form a composite object of informational relevance.* <http://www.oclc.org/research/projects/archive/wcp/pubs/rn1-websites.htm>

¹⁷¹ O'Neill, E.T. Characteristics of web accessible information. // IFLA Journal, 24, 2(1998), str. 114-116.

¹⁷² Adrian Brown upotrebljava termin statična i dinamična mrežna mjesta. Statična su sastavljena iz serije stranica, koje su povezane poveznicama, a svaka je od njih sastavljena iz datoteka koje su hijerarhijski smještene na poslužitelju. Njihova je lokacija vidljiva iz URL-a. Za razliku od statičnih, dinamična mrežna mjesta generiraju se nakon prihvaćanja zahtjeva klijenta i omogućavaju pristup malom dijelu sadržaja. U tu se skupinu ubrajaju: baze podataka, sastavljeni sadržaji iz različitih poslužitelja, skripte i različite personalizacije sučelja. Vidi: Brown, A. Archiving websites : a practical guide for information management professionals. London: Facet Publishing, 2006. Str. 44-45.

¹⁷³ Steenbakkera, J. Setting up a Deposit System for Electronic Publications: The NEDLIB Guidelines. The Hague : Koninklijke Bibliotheek, 2000. Str. 7.

Steenbakk¹⁷⁴ razlikuje mrežne publikacije i po njihovu izvoru nastanka. Tako postoje publikacije koje su izvorno nastale u elektroničkom obliku i publikacije koje su posljedica neke reprodukcije, odnosno digitalizacije iz različitih medija.

Među prvim autorima koji su pokušali uvesti kategorizaciju publikacija na mreži je O'Neill,¹⁷⁵ koji dijeli mrežnu građu na mrežne stranice i mrežna mjesta, zatim mrežne stranice svrstava po sadržaju i funkciji u sljedeće kategorije:

- **nebeletristična literatura** (*non-fiction*): izvještaji, znanstveni doprinosi, članci, komentari, eseji, uvodnici, knjige;
- **beletristika** (*fiction*): humor, lijepa književnost, igrice, materijal za zabavu;
- **referentne stranice/indeksi**: naslovnice, indeksi, izvadci, sažetci, stranice s podacima (*fact sheets*), direktoriji, profili, bibliografije, vodiči, biografske i autobiografske informacije, poprečne unakrsne reference (*cross-references*) do vanjskih izvora;
- **institucionalne**: mrežna prodaja, oglasi, informacije o proizvodima, prodajni katalozi, priručnici, promocijska građa, institucionalni opisi;
- **osobne**: stranice koje prikazuju informacije o pojedincima ili skupinama.

U svojoj podjeli O'Neill ne uzima u obzir strukturu i oblik različitih kategorija. Isključuje sve stranice o mrežnim poslužiteljima. Također izjednačava baze podataka s pretraživačkim stranicama ili računalne igrice s beletristikom, iako imaju vrlo različitu funkciju.

S druge strane, O'Neill dijeli mrežna mjesta po načinu pristupa:

- **javno mrežno mjesto** - omogućava slobodan pristup njegovim sadržajima;
- **privatno mrežno mjesto** - sprečava pristup korisnicima koji nemaju dopuštenje, sadržaji nisu namijenjeni širokoj javnosti;
- **neaktivna mrežna mjesta** – to su šablone (*templates*), odnosno stranice koje obavještavaju da je sadržaj u izgradnji, ili stranice koje nemaju korisnih sadržaja i na njima je minimalna količina informacija;

¹⁷⁴ Steenbakk, J. Idem.

¹⁷⁵ O'Neill, E. T.: Characteristics of web accessible information. Presented at 63rd IFLA General Conference, August 31-September 5, 1997. // IFLA journal. 24, 2(1998) <http://wotan.liu.edu/doi/data/Articles/julksrnqay:1998:v:24:i:2:p:7224.html>. (15.12.2006.), str. 115.

- **indeksi** – to su direktoriji datoteka za mrežno mjesto; dostupni su pomoću pretraživača (preglednika, *browser*) u slučajevima kada nema stalne stranice za to mrežno mjesto; obično obavještavaju da je stranica neaktivna ili nije spremna za gledanje.

Podjela mrežnih mjesta s obzirom na njihov fokus.¹⁷⁶

- **opisne mrežne stranice** (deskriptivne): omogućavaju informacije o sponzoru mrežnoga mjesta, koji može biti osoba ili organizacija;
- **uslužne mrežne stranice**: njihov je cilj izvođenje neke uslužne djelatnosti, više nego li obavještavanje o sponzoru.

S obzirom na to da su O'Neillove podjele nastale vrlo rano, u njegovim se pokušajima kategorizacije mrežne građe opaža pokušaj što širega obuhvaćanja novih oblika i inzistiranje na analogiji tradicionalnih oblika građe. Zato i opažamo velike nedosljednosti u tim podjelama.

Velik doprinos u formalnoj bibliografskoj podjeli i definiciji nove građe na mreži je revizija pravilnika AACR,¹⁷⁷ koji je u svom 9. i 12. poglavlju uveo nove tipove publikacija: integrirajuću i neomeđenu građu. To je veoma općenita podjela, jer se u njoj gubi bogatstvo svih prije popisanih oblika. Prvi međunarodni standard koji je prihvatio klasifikaciju građe po AACR2 je ISBD(CR).¹⁷⁸ U pripremi je i revizija ISBD(ER), koja će nam pomoći kod katalogizacije tih izvora.

Novi ISBD (ISBD(CR))¹⁷⁹ razlikuje mrežnu stranicu od mrežnog mjesta i uvodi koncept neomeđene i integrirajuće građe.

¹⁷⁶ Po predavanju O'Neilla na IFLA 1997. Ti podatci nisu objavljeni u gore spomenutom članku, iako je autor na samom predavanju distribuirao prezentaciju u PPT s navedenim informacijama.

¹⁷⁷ Revising AACR2 to ACcommodate Seriality, 1999: Report to the Joint Steering Committee for Revision of AACR, April 1999. URL: <http://www.collectionscanada.ca/jsc/ser-rep4.html>. (3.11.2006.)

¹⁷⁸ ISBD(CR) : International standard bibliographic description for serials and other continuing resources : revised from ISBD(S) : international standard bibliographic description for serials / recommended by the ISBD(S) Working Group ; approved by the Standing Committees of the IFLA section on Cataloguing and the IFLA Section on Serial Publications. München : Saur, 2002. - XI, 112 str.. (UBCIM publications. New series, ISSN 0941-8946 ; Vol. 24).

¹⁷⁹ ISBD(CR) : međunarodni standardni bibliografski opis serijskih publikacija i druge neomeđene građe : prerađeno izdanje ISBD(S): Međunarodnoga standardnoga bibliografskog opisa serijskih publikacija / [priređivač] Međunarodni savez knjižničarskih društava i ustanova ; [s engleskog prevela Ana Barbarić]. Zagreb : Hrvatsko knjižničarsko društvo, 2005. 137 str.. - (Povremena izdanja Hrvatskog knjižničarskog društva, ISSN 1334-7667. Novi niz ; knj.11).

Neomeđena građa (*continuing resources*) je "Jedinica građe koja se izdaje tijekom nekog razdoblja bez unaprijed utvrđenog kraja izlaženja. Neomeđena građa obuhvaća serijske publikacije i tekuću integrirajuću građu",^{180,181} dok je **integrirajuća građa** (*integrating resources*) "Jedinica građe koja se nadopunjuje ili mijenja osuvremenjivanjima koja ne ostaju zasebna, već su uklopljena u cjelinu. Integrirajuća građa može biti omeđena ili neomeđena. Primjeri za integrirajuću građu obuhvaćaju građu koja se osuvremenjuje poput publikacija koje imaju uvez sa slobodnim listovima i mrežnih mjesta."^{182,183}

5.2. Podjele mrežne građe na osnovi njihova sadržaja

Osim spomenutih formalnih podjela mrežne građe, kod raznih autora nalazimo tipologije mrežne građe koje zalaze u sadržaj, informacijsku funkciju, formalni oblik ili u njihovu kombinaciju kroz definiciju novih žanrova. U nastavku su prikazane neke od tih podjela.

Kod Haberta, Hurault-Planteta i Jardina¹⁸⁴ nalazimo podjelu mrežnih stranica po njihovoj funkciji. U okviru jednog istraživačkog projekta koristili su se klasifikacijom poduzeća Nielsen/NetRatings, koji navodi 70 kategorija za različite grane djelatnosti. Partneri su, na osnovi tih kategorija polja, empirijski odredili skupinu sljedećih dvanaest kategorija:

1. Stranica dobrodošlice (*Welcoming page*) koja obavještava o tematici mrežnoga mjesta
2. Indeks ili kazalo sadržaja (*Headings index*) za hijerarhijsku navigaciju kroz mrežno mjesto.
3. Sažetak (*Summary*) za slobodnu navigaciju po mrežnom mjestu
4. Vanjske poveznice (*External links*) za navigaciju po mreži

¹⁸⁰ *Continuing resource: A bibliographic resource that is issued over time with no predetermined conclusion. Continuing resources include serials and ongoing integrating resources.* (ISBD(CR) str. 4)

¹⁸¹ ISBD(CR), hrvatski prijevod, str. 18.

¹⁸² ISBD(CR), hrvatski prijevod, str. 16.

¹⁸³ *Integrating resource: A bibliographic resource that is added to or changed by means of updates that do not remain discrete and are integrated into the whole. Integrating resources can be finite or continuing. Examples of integrating resources include updating loose-leafs and updating Web sites.* (str. 5).

¹⁸⁴ Habert, B. ; Hurault-Plantet, M. ; Jardina, M. Web page classification: field categories and functional categories. SENSNET RNRT Project, 2005. URL: <http://www.limsi.fr/RS2005/chm/lir/lir4/>. (15.12.2006.)

5. Komunikacijske stranice (*Communication*) za izmjenu informacija
6. Stranice za kontakte (*Contact*) za uspostavljanje kontakta s vlasnikom mrežnih stranica.
7. Stranice sa sadržajima (*Content*) za obavještanje o određenoj temi
8. Katalozi ili direktoriji (*Catalogue, Directory*) za pridobivanje dodatnih informacija o određenoj temi
9. Stranice sa sučeljem za informacijsko pretraživanje (*Request form for information retrieval*) za plasiranje informacijskog zahtjeva
10. Stranice sa sučeljem s odgovorom na informacijsko pretraživanje. (*Answering request for information retrieval*)
11. Stranice za prijavu (*Login*), odnosno sučelje za pristup zaštićenom mrežnom mjestu
12. Stranice koje obavještavaju o pogreškama ili nedostupnim stranicama (*Error*).

Prime-Claverie, Beigbeder i Lafouge¹⁸⁵ tvrde da možemo povećati preciznost pretraživanja dokumenata po internetu pomoću dodatnih metapodataka, koji se odnose na tipologiju mrežnih mjesta i mrežnih stranica. Njihova se tipologija mrežnih mjesta temelji na tri aspekta:

- Mrežna mjesta po tipu odgovornog tijela/stvaratelja: mrežna mjesta institucija, poduzeća, društava i pojedinaca.
- Mrežna mjesta po informacijskoj funkciji:
 - **Dučanski prozori** (*shop-window site*) - sadrže opisne informacije osoba ili organizacija koje stvaraju te informacije (aktivnosti, proizvodi, partneri, kontaktni naslovi...) u tu skupinu ulaze naslovnice.
 - **Mjesta za traženje i pretraživanje** - omogućavaju pristup izvoru informacija na mreži.
 - **Mrežna mjesta s izvorima informacija** - organiziraju i nude svoje vlastite izvore informacija.
 - **Mrežne usluge** - nude usluge koje su vezane uz život na mreži i internetu, kao što su e-pošta, novi forumi itd.
- Mrežna mjesta po tipovima sadržanih informacija:

¹⁸⁵ Prime-Claverie, C. ; Beigbeder, M. ; Lafouge, T. Transposition of the cocitation method with a view to classifying web pages. // Journal of the American Society for Information Science and Technology. 55, 14(2004), str.1282-1289.

- 'Samoopisne' informacije, koje se odnose na stvoritelja stranica;
- 'Nesamoopisne' informacije.

Asirvatham i Ravi¹⁸⁶ su pomoću statističkih algoritama i algoritama za strojno učenje istražili mogućnost kategorizacije mrežnih stranica kako bi poboljšali indeksiranje i povećali preciznost njihova pretraživanja. Njihovo se istraživanje temelji na teoriji vjerojatnosti i frekvencijske distribucije riječi. Osim metode za kategorizaciju mrežnih stranica koja se temelji na sadržaju, obrađuju i metode za strukturnu kategorizaciju. Pritom se uzimaju u obzir lokacije poveznica (*links*), tekstovi i slike (uključujući formule i grafikone). Te se informacije lako mogu dobiti iz dokumenta u html. Ustanovili su da mogu automatski podijeliti mrežne stranice u tri opće kategorije:

- **informativne stranice** - sadrže logo na vrhu, prati ih navigacijska palica (*navigation bar*) koja ima poveznice na druge važne stranice; kod tih su stranica česte poveznice na normalne tekstove; paleta boja na slikama je veća; imaju veći broj prirodnih slika (fotografija);
- **istraživačke stranice** - obično imaju velik broj tekstova, formula i grafova u obliku slika itd. Imaju veći broj umjetnih slika koje prikazuju grafove i formule; često imaju binarne slike;
- **osobne naslovnice** - sadrže ime, adresu i sliku osobe koja se običajno nalazi na vrhu stranice; poveznice vode do publikacija osobe, ako ih ima i na njezine omiljene lokacije na mreži.

Automatskom kategorizacijom postigli su 87,83% uspješnosti.

Na mrežnim stranicama akademskih ustanova često nalazimo podjele elektroničkih izvora informacija po njihovoj funkciji i po sadržaju. Naprimjer, Sveučilišna knjižnica u Chicagu¹⁸⁷ dijeli elektroničke publikacije u sljedeće kategorije:

- istraživački vodiči - sadrže relevantne informacije u vezi s određenim područjem; organizirani su po predmetnom kazalu i u njima se mogu naći osobe za kontakt iz tog područja te odgovarajuće informacije o tiskanim i elektroničkim zbirkama u knjižnici, kao i poveznice na relevantna mjesta na mreži;

¹⁸⁶ Asirvatham, A. P. ; Ravi, K. K. Web page categorization based on document structure, 2001. URL: <http://1.1.1.1/269697092/259139976T060811105245.txt.binXMysM0dapplication/pdfXsysM0dhttp://gdit.iiit.net/~ar ul/paper.pdf>. (15.12.2006.)

¹⁸⁷ URL: <http://www.lib.uchicago.edu/e/net/types.html>. (15.12.2006.)

- indeksi bibliografskih podataka (običajno imena autora, naslova ili ključnih riječi) o člancima u znanstvenim i stručnim revijama, istraživačkim i akademskim radovima, zbornicima i slično;
- elektroničke knjige i tekstovi;
- elektroničke revije;
- knjižnični katalozi - lokalni i skupni;
- referentna literatura, koja obuhvaća elektroničke enciklopedije, rječnike, almanahe i druge referentne publikacije;
- statistički izvori informacija;
- popis zvučnih snimaka;
- baza podataka slikovnih zbirka (iz područja umjetnosti, kartografske građe, medicine itd.).

Godine 1992. Yates i Orlikowski¹⁸⁸ uvode novu teorijsku disciplinu koja publikacije na mreži tretira kao komunikacijske jedinice koje sa svojim zajedničkim svojstvima tvore žanrove. Nazivaju je genologija interneta ili žanrovi na internetu. Tradicionalno su "žanrovi literarni i retorički koncepti koji opisuju opće prihvaćene tipove govora ili komuniciranja (primjerice roman ili prodika). U kontekstu organizacijskog komuniciranja može se taj koncept koristiti za priznate tipove komuniciranja (primjerice pisma, interna saopćenja ili sastanci) koji se karakteriziraju strukturalnim, lingvističkim i sadržajnim konvencijama".¹⁸⁹ Zanimljivo je da kad se ti žanrovi prenesu u drugi medij, ljudi se njima koriste, odbace ih ili prilagode svojim potrebama.^{190,191} Moguće je da se određena vrsta sjedini s drugom ili nastane sasvim nov žanr na koji dosad još nismo naišli u medijima. Često su to jedinstveni primjeri, iako se događa da neke od tih žanrova prihvati skupina ljudi i oni tako ožive kao standardni oblik komuniciranja.

¹⁸⁸ Yates, J. ; Orlikowski, W. J. Genres of organizational communication: a structural approach to studying communication and media. // *The Academy of Management Review*. 17, 2(1992), str. 299-326.

¹⁸⁹ Yates, J. ; Orlikowski, W. J., idem: "Genre is a literary and rhetorical concept that describes widely recognized types of discourse (e.g., the novel, the sermon). In the context of organizational communication, it may be applied to recognized types of communication (e.g., letters, memoranda, or meetings) characterized by structural, linguistic, and substantive conventions."

¹⁹⁰ Orlikowski, W.J. ; Yates, J. Genre repertoire: the structuring of communicative practices in organizations. // *Administrative Sciences Quarterly* 33(1994), str. 541-574.

¹⁹¹ Po riječima Orlikowske i Yatesa "Ljudi stvaraju, reproduciraju i mijenjaju žanrove kroz proces strukturiranja". (str. 545); Shepherd i Wattersova tvrde da "Funkcionalnost, koju omogućava novi medij, vodi evoluciju podvojenih žanrova naime onih koji su nastali na osnovi već postojećih žanrova u drugim medijima sa njihovim varijacijama, dok ne nastaju novi, sasvim različiti od izvornih." Vidi djelo: Shepherd, M. ; Watters, C. The functionality attribute of cybergenres. // *Proceedings of the 32nd Hawaii International Conference on System Sciences (HICSS'99)*, 1999. URL: <http://csdl2.computer.org/comp/proceedings/hicss/1999/0001/02/00012007.PDF>. (15.12.2006.)

Naprimjer, tradicionalna knjiga ili znanstveni članak preselili su se na internet bez promjena. Elektroničke revije još su u tranziciji. Način citiranja također se mijenja jer upravo digitalno okruženje omogućava da iz reference odmah pristupamo do cijeloga teksta.

Shepherd i Watters¹⁹² uvode termin *cybergenre* (kiberžanr) u značenju žanrova na internetu. Po njima ne određuju samo njihov sadržaj i format, koji su specifični za tradicionalne žanrove, već i funkcionalnost, koja je specifična za digitalno okruženje. U okviru njihova istraživanja na osnovi spomenutih svojstava klasificirali su mrežne stranice u 6 skupina (*cybergenres*):

- Naslovnica
- Brošura
- Izvor (*resource*)
- Katalozi
- Pretraživači
- Računalne igrice.

Znanstvena istraživanja na području genologije interneta dovela su do definicije više žanrova koji su nastali u zadnjih nekoliko godina i koje možemo definirati kao određen tip publikacije.

U okviru različitih studija na svjetskoj mreži mogu se naći definicije sljedećih žanrova:

- **FAQ - *Frequently Asked Questions*** - preko Alta Viste 1996. godine ustanovili su da postoji 170,000 mrežnih stranica sa FAQ u naslovu.¹⁹³
- **Naslovnica** (osobna ali organizacijska) - postoje različite teorije o njenom nastanku; Joanne Yates¹⁹⁴ tvrdi da su naslovnice vjerojatno nastale adaptacijom Unixovih *.plan datoteka, koje su se tiskale kada je korisnik zahtijevao status svog računa (*account*); organizacijske mrežne

¹⁹² Shepherd, M. ; Watters, C. The functionality attribute of cybergenres. // Proceedings of the 32nd Hawaii International Conference on System Sciences (HICSS'99), 1999. URL: <http://csdl2.computer.org/comp/proceedings/hicss/1999/0001/02/00012007.PDF>. (15.12.2006.)

¹⁹³ Crowston, K. ; Williams, M. Reproduced and emergent genres of communication on the world wide web. // The Information Society, 16 (2000), str. 207-208.

¹⁹⁴ Crowston, K. ; Williams, M. (2000) Idem, str. 207.

stranice vjerojatno su nastale po uzoru na opis profila predavača i fakulteta; naslovnica je po njima *'page d'accueil'* i obavlja dvije velike funkcije: prvo, uvodi korisnika u opći sadržaj mrežnog mjesta s informativnim kazalima sadržaja, i drugo, djeluje kao formalni portal u mrežno mjesto i čitatelju omogućava pristup i navigaciju po njezinim mrežnim sadržajima; Askehave i Nielsen¹⁹⁵ tvrde da sama specifičnost naslovnica određuje tu vrstu kao novu i kao poseban žanr.

- **Informacijski portali** - razlikuju se od naslovnica po većem broju izlaznih poveznica; njihov je cilj organizirati izvore informacija po različitim kriterijima i omogućiti lakši pristup.
- **Hotlists**¹⁹⁶ su mrežne stranice s vanjskim poveznicama do građe nad kojom kreator nema nikakve kontrole.
- **Topical home pages**¹⁹⁷ raširene su hotlists s dodatnim informacijama.
- **Mrežne stranice o mrežnim poslužiteljima**¹⁹⁸ obavještavaju o djelovanju mrežnih poslužitelja i sasvim su nov žanr. Primjer takvih stranica su:
 - **Under construction** (stranica je u izgradnji) - obično sa slikom radnika koji simbolizira čovjeka u radu;
 - **Custom 404** (Pravilo 404) - standardne stranice koje obavještavaju da pretraživač ne nalazi tražene stranice, a sadrže informacije ponuđača usluga i njegov logotip;
 - **Web site has moved** (stranica je na drugoj lokaciji) - obavještava nas o novom URL-u za traženu mrežnu stranicu;
 - **Popis direktorija datoteka** - popis datoteka koje su dostupne na određenom direktoriju ili na mrežnom poslužitelju, primjerice u obliku FTP direktorija;

¹⁹⁵ Askehave, I. ; Nielsen, A. E. What are the characteristics of digital genres? - Genre theory from a multi-modal perspective // Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS'05) - Track 4, 2005. URL: <http://csdl2.computer.org/persagen/DLAbsToc.jsp?resourcePath=/dl/proceedings/&toc=comp/proceedings/hicss/2005/2268/04/2268toc.xml>. (15.12.2006.), str. 98a.

¹⁹⁶ Crowston, K. ; Williams, M. op.cit., str.208.

¹⁹⁷ Crowston, K. ; Williams, M. op.cit, str. 208.

¹⁹⁸ Crowston, K. ; Williams, M. op.cit., str. 209.

- **Web server statistics** (statistika mrežnoga poslužitelja) - sadrži statističke podatke, kao što su broj rezultata pretraživanja, posjeta mrežne stranice ili mrežnoga mjesta itd.
- **Interaktivne mrežne stranice**¹⁹⁹ - obično su to sučelja različitih baza podataka u koje unosimo podatke ili na osnovi traženih parametara dobivamo određen rezultat pretraživanja (primjerice vozni redovi, telefonski imenici, statistički podatci, astrološke baze itd.);
- **Blogovi**²⁰⁰ - to su mrežne stranice koje se često mijenjaju jer se tekstovi dodaju po vremenskom redoslijedu, tako da je na udaru uvijek najnoviji upis;^{201,202} blogove često izrađuje i održava jedna osoba, sadržaj je odraz njezinih interesa i zanimanja; zamjenjuju naslovnice osoba;²⁰³ posebnost blogova je u tome da imaju arhive (poveznice do starijih stranica) i značke (male sličice u kutu s oglasima).

¹⁹⁹ Crowston, K. ; Williams, M. op. cit., str. 209.

²⁰⁰ Herring, S. C. ... [et al.]. Bridging the gap : a genre analysis of weblogs // Proceedings of the 37th Hawaii International Conference on System Sciences (HICSS'04) - Track 4, 2004. URL: <http://1.1.1.1/260056852/282918808T060814112148.txt.binXMysM0dapplication/pdfXsysM0dhttp://www.ics.uci.edu/~jpd/classes/ics234cw04/herring.pdf>. (15.12.2006.), str. 4010.

²⁰¹ Winer, D. The history of weblogs, 2002. URL: <http://oldweblogscomblog.scripting.xom/historyOfWeblogs>. (15.12.2006.)

²⁰² Winer navodi da je Tim Berners-Lee izradio prvi weblog u CERN-u. Herring, ... [et al.] tvrde da je Jorn Barger prvi put upotrijebio izraz "weblog". Današnju izvedenicu "blog" stvorio je Peter Merholz početkom 1999. godine, kada je "weblog" počeo izgovarati kao "wee-blog".

²⁰³ Blood, R. u svome djelu The Weblog handbook : practical advice on creating and maintaining your blog, Cambridge MA : Perseus Publishing, 2002, razlikuje tri osnovna tipa weblogova: filteri, kod kojih je sadržaj iz vanjskog okruženja autora bloga (naprimjer svjetski događaji, događaji na mreži itd.), osobni žurnali, kod kojih blogger iznosi svoja gledišta i stajališta o određenoj tematici, i zapisi (*notebooks*), koji imaju oblik dužih i fokusiranih eseja. Krishnamurthy, S. u djelu The Multidimensionality of blog conversations: the virtual enactment of September 11. Maastricht, The Netherlands: Internet Research 3.0, 2002, predlaže klasifikaciju blogova na četiri tipa s obzirom na dvije različite dimenzije tj.: osobna (*personal*) nasuprot tematskim i individualna nasuprot skupnim.

- **Baze podataka** - na mreži vidimo samo sučelje za pristup, u pozadini je program koji omogućava smještanje i pristup podacima. Ashley²⁰⁴ razlikuje 4 tipa baza podataka: otvorene (nisu zaključene i još se dopunjuju), zatvorene (zaključene su), statične (kod njih je moguć direktan uvid u podatke i ne mijenjaju se) i dinamične (u njima se informacija kod svakog zahtjeva nanovo stvara).²⁰⁵
- **Mrežne enciklopedije**²⁰⁶ - enciklopedija je djelo koje sadrži informacije o svim znanstvenim područjima, odnosno detaljno pokriva sve grane znanosti; sastoji se od članaka i indeksa;²⁰⁷ enciklopedije se pojavljuju u različitim oblicima na mreži; na naslovnici je sučelje koje nam omogućava pretraživanje; na zahtjev se pojavi stranica a objašnjenjem i opisom sadržaja traženog pojma; primjeri najpoznatijih mrežnih enciklopedija su Wikipedija²⁰⁸ i Everything2. Stvaraju ih korisnici sami.
- **Mrežni rječnici** - stvaraju se slično kao i mrežne enciklopedije.
- **Word-Of-Mouth (WOM)** – to je žanr koji je nedavno prenesen iz marketinga, a otkrio ga je Pollach,²⁰⁹ to je neformalna potrošačka ocjena o određenim proizvodima, markama (*trade mark*) ili uslugama; nastao je iz usmenih saopćenja i stekao priznanje mnogih korisnika interneta; danas ima svoju bitnu strukturu koja se sastoji od tri dijela: ocjene dobrih svojstava potrošačkoga proizvoda, ocjene loših svojstava te opće ocjene; bitan je i jezik, kao i dužina

²⁰⁴ Ashley, K. The preservation of databases. // VINE : The Journal of Information and Knowledge Management Systems. 34, 2,(2004), str. 66-70. <http://docserver.emeraldinsight.com/deliver/cw/mcb/03055728/s3/p66.htm?fmt>. (24.2.2005.)

²⁰⁵ Ashley navodi kao primjer otvorene-statične baze podataka popis transakcija na tekućem računu; bilanca tekućeg računa primjer je otvorene-dinamične baze. Statistički podatci iz prošlih popisa stanovništva primjer su zatvorene-statične baze, dok je popis bilanca mrtve banke (to je banka, koja je prestala posloovati, zatvorena je i ne djeluje više) primjer zatvorene-dinamične baze podataka.

²⁰⁶ Emigh, W. ; Herring, S. C. Collaborative authoring on the web: a genre analysis of online encyclopedias. // Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS'05) - Track 4, 2005. <http://1.1.1.1/270600412/290040792T060814112453.txt.binXMysM0dapplication/pdfXsysM0dhttp://csdl2.computer.org/comp/proceedings/hicss/2005/2268/04/22680099a.pdf> (15.12.2006.), str. 99a.

²⁰⁷ Merriam-Webster Online Dictionary URL: <http://www.m-w.com/cgi-bin/dictionary?book=Dictionary&va=encyclopedia&x=18&y=9>. (9.09.2005.)

²⁰⁸ Wiki je komunikacijski mehanizam skupina koji je 1995. godine izmislio Ward Cunningham, tako da svi korisnici mogu stvarati i popravljati sadržaj mrežnih stranica bez ikakvih ograničenja.

²⁰⁹ Pollach, I. Electronic word of mouth: a genre analysis of product reviews on consumer opinion web sites. // Proceedings of the 39th Hawaii International Conference on System Sciences (HICSS'06) - Track 3, 2006. <http://1.1.1.1/343626388/272137528T060814112646.txt.binXMysM0dapplication/pdfXsysM0dhttp://csdl2.computer.org/comp/proceedings/hicss/2006/2507/03/250730051c.pdf> (15.12.2006.), str.51c.

rečenica, leksičko bogatstvo, negacije, lične zamjenice, formalnosti i paralingvističke posebnosti.

- **Bibliografije na mreži** - Hendry i Carlyle²¹⁰ tvrde da je iz tradicionalnih bibliografija nastalo mnogo novih žanrova na mreži, kao što su hotlists, blogovi, popisi referenca, stranice s popisima poveznica do građe koja je relevantna za određeno područje, direktoriji, portali, digitalne knjižnice i slično; mrežno je okruženje najviše pridonijelo razvoju bibliografijama tako što se iz statičnoga popisa bibliografskih referenca u tiskanom mediju prešlo na popis poveznica do cijelih tekstova; Hendry i Carlyle čak tvrde da se tiskana bibliografija razvila u digitalnu zbirku; razlika je i u načinu njihova nastanka; dok u tiskanom mediju bibliografije obično stvaraju pojedinci, u digitalnom okruženju one su rezultat rada skupine ljudi. Postoje različiti tipovi bibliografija na mreži:²¹¹ otvorena, koju može bilo tko stvarati, vođena (*moderated*), s recenzijom (*peer-reviewed*) ili miješana.
- **RFC** (*Request for Comments*) - oblik komuniciranja među informatičarima o standardizaciji postupaka ili specifikacijama određenoga rješenja vezanoga uz djelovanje interneta ili njegovih usluga; RFC-e često računalna javnost prihvaća kao norme ili standarde.

Pregled različitih teorija i podjela publikacija na mreži pokazuje da je teško odrediti opću klasifikaciju tih publikacija. Problem je utoliko veći što još uvijek razmišljamo u dimenzijama tiskane građe, tj. analogno, a ne digitalno. Problematika određivanja različitih tipova mrežne građe vrlo je složena i ako bismo željeli obuhvatiti svu građu, morali bismo primijeniti više kriterija. Upravo se genealogija interneta temelji na sadržaju, strukturi, obliku i funkciji elektroničke građe i pomaže kod identifikacije novih oblika informacija i građe, koji dobivaju svoj identitet na mreži. Ti novi oblici ili žanrovi sve se više razlikuju od klasičnih oblika informacija, a zbog vrlo snažne tradicije vrlo polako prodiru u bibliotekarstvo. Velik doprinos priznavanju njihova postojanja u bibliotekarstvu jest uvođenje koncepta integrirajuće i

²¹⁰ Hendry, D. G. ; Carlyle, A. Hotlist or bibliography? A case of genre on the web. // Proceedings of the 39th Hawaii International Conference on System Sciences (HICSS'06) - Track 3, 2006. <http://1.1.1.1/273447484/39443704T060814112605.txt.binXMysM0dapplication/pdfXsysM0dhttp://csdl2.computer.org/comp/proceedings/hicss/2006/2507/03/250730051b.pdf>. (15.12.2006.), p. 51b.

²¹¹ Henry, D. G. ; Carlyle, A. Idem.

neomeđene građe. Međutim, ova su dva koncepta previše općenita i mnogim katalogizatorima još nije sasvim jasno njihovo razgraničenje. U samom uvodu ISBD(CR) čak se i napominje da se sva integrirajuća građa jednako obrađuje, bez obzira na to je li omeđena ili neomeđena. To je još jedan dokaz da se formalna bibliografska obradba ne upušta u funkciju, strukturu ili sadržaj elektroničke građe. Ta činjenica može produžiti vrijeme potrebno da se prihvate novi elektronički oblici, za koje je nemoguće naći analogiju u klasičnim oblicima građe.

Bibliotekarska znanost morala bi uvesti novu tipologiju mrežne građe, koja bi uzela u obzir njihove nove strukture, funkcionalnost i sadržaje. Pritom bi se mogla koristiti dostignućima genologije interneta. Zašto je to toliko važno? Bibliotekari već dugo nisu samo kustosi knjižničnih fondova, već su organizatori znanja i informacijski stručnjaci. Njihov je zadatak da korisnicima pomažu pri savladavanju informacijske zasićenosti, pri probijanju kroz šumu relevantnih i nerelevantnih informacija te konačnom nalaženju traženih informacija. Tipologija elektroničke građe svakako bi pomogla dodatnoj organizaciji informacija, čiji je glavni cilj uspješnije pretraživanje. Trenutačno, s postojećim internetskim pretraživačima, vrlo teško nalazimo relevantne informacije na mreži. U budućnosti će to biti još teže. Zbog prevelike količine mrežne građe, knjižničari već sada nisu u mogućnosti da katalogiziraju svu relevantnu građu koja je objavljena na svjetskoj mreži. Osim toga, postoji i pitanje vjerodostojnosti same građe. Razumljivo je da mrežna građa dobiva svoje karakteristike, a rezultat je rada njezinih autora. Možemo li bibliotekari utjecati na formaliziranje tih oblika? Budemo li definirali oblike, koje uključujemo u proces obradbe, vjerojatno ćemo utjecati na njihovo kreiranje. Osim toga, s obzirom na rast mrežne građe, stvaranje nove tipologije mrežne građe omogućilo bolji bibliografski opis, a time bi i pretraživanje po toj građi u digitalnim arhivima bilo preciznije. Njihovo formalno postojanje utjecalo bi na uvođenje preciznije bibliografske obrade.

Zbog strogih standarda i pravilnika u bibliotekarskoj znanosti novi bi oblici mrežne građe mogli biti prije prihvaćeni u dokumentalistici, odnosno sadržajnoj obradbi, nego li u formalnoj obradbi.

6. KRITERIJI ZA ODABIR ELEKTRONIČKE GRAĐE

6.1. Definicija kriterija odabira mrežne građe

Prvi pokusi preuzimanja mrežne građe temeljili su se na kriteriju nacionalnih domena, što je logično jer najviše nacionalne domene (*.si, *.hr) dodjeljuju domaće firme i označuju geografsko područje. Nacionalna domena se zato obično dodjeljuje javnim institucijama, među njima su istraživačke ustanove i samostalni istraživači. Privatne i komercijalne firme imaju druge domene, kao što su *.org, *.com, *.net i slično. Međutim, zbog različitih cijena dodjeljivanja i gostovanja domena ili zbog mogućnosti stvaranja igre riječima s domenskim nastavkom mnogi pojedinci ili firme kupuju domene u drugim zemljama. Budući da su to više iznimke nego pravilo, s prikupljanjem mrežne građe u okviru najviših nacionalnih domena možemo relativno jednostavno i brzo prikupiti veći dio javno dostupne domaće mrežne građe. Kod manjih naroda, kao što je u Sloveniji ili Estoniji, mogu se uključiti dodatne domene privatnih poduzeća i ograničiti njihovo prikupljanje leksičkim sredstvima, kao što su različite sintaktičke ili morfološke strukture. Te mogućnosti ovise i o sposobnostima odnosno funkcijama robota za prikupljanje mrežne građe.

Kod automatskoga prikupljanja mreže na prvi pogled izgleda da nije prisutna nikakva selekcija. Zato taj pristup pogrešno nazivaju neselektivnim. Međutim, u tom je pristupu i te kako potrebno definirati pravila selekcije. Ona se izvode u fazi pripreme ili namještanja alata za prikupljanje građe. Obično većina robota ima mogućnost namještanja početka prikupljanja na odabran popis URL-a, na najviše domene, na razinu dubine prikupljanja ili stranica po mrežnom mjestu ili na veličinu datoteka u MB. Sofisticiraniji roboti imaju i druge mogućnosti, noviji imaju čak i sposobnost prepoznavanja jezika. Ovaj je pristup neselektivan, što znači da nije potrebna neka veća intervencija nad sadržajem građe, drugim riječima prikupljena je sva građa, i ona koja je iz sadašnje perspektive nerelevantna.

Glavni argument za automatsko prikupljanje mrežne građe je činjenica da u trenutku skupljanja nikako ne možemo točno znati koji će sadržaji biti vrijedni za buduće generacije. Već se danas vidi da su neki od sadržaja, koji su u trenutku prikupljanja smatrani beznačajnom građom, postali

predmet različitih suvremenih istraživanja o internetu.²¹² Istina je da automatsko skupljanje mrežne građe ne jamči optimalnu kvalitetu jer je nemoguće pregledati je li prikupljanje milijuna URL-a bio sasvim uspješan i jesu li sva mrežna mjesta u cjelini zahvaćena. Ovakva bi nas kontrola koštala više od strogoga selektivnoga prikupljanja. Međutim, postoji i prednost automatskoga prikupljanja, a to je činjenica da mrežna građa nije izvučena iz konteksta kao kod ručnoga prikupljanja pa vanjske poveznice djeluju. Zato je i informacija, koja se tim putem sačuva, potpunija.

Prva nacionalna knjižnica koja je primijenila vrlo stroge kriterije selekcije je australska. U dvanaest godina razvili su vrlo bogatu praksu, koja je dokumentirana u pravilnicima i predstavljena na stranicama Pandore u okviru integralne politike i strategije gradnje zbirke. U tom razdoblju uspjeli su prikupiti više od dvanaest tisuća naslova. Razlog za selektivno prikupljanje je u prijenosu prakse iz tradicionalne materijalne građe u elektroničko okruženje. Jedinice građe brižljivo su prikupljene, uspoređene s tradicionalnim oblicima (knjige, serijske publikacije i slično), a zatim klasično bibliografski obrađene te smještene u katalog i u arhiv. Takva je praksa dugotrajnija i složenija nego kod tradicionalnih oblika, jer zahtijeva provjeravanje kvalitete, unos dodatnih metapodatka koji ovise o formatu i obliku elektroničke građe. Australsku praksu i ručno prikupljanje mreže usvojile su zatim i Kongresna knjižnica u SAD-u, Britanska knjižnica, Nacionalna knjižnica Kanade i ostale, iako su koristile tematski pristup. Tematsko prikupljanje u spomenutim zemljama manje je zahtjevno od australske prakse, jer često obrađuju više mrežnih mjesta jednim bibliografskim opisom.

Budući da obje metode, automatska i ručna metoda prikupljanja građe, imaju svoje prednosti i nedostatke, mnoge nacionalne knjižnice analiziraju mogućnost upotrebe obaju pristupa. U Nacionalnoj knjižnici, Australije naprimjer, svjesni su da se tako podrobnim odabirom i obradom tek reprezentativnoga dijela mreže zanemaruje veći dio, koji bi u budućnosti mogao biti relevantan za znanstvenike i istraživače. S druge pak strane, naprimjer u Francuskoj nacionalnoj knjižnici, primjećuju da se nekim mrežnim mjestima potrebno posebno posvetiti, jer su dio dubinske mreže, ili ih roboti prikupljači ne vide, ili se mijenjaju brže nego li ih robot uspijeva

²¹² Thelwall, M. ; Vaughan, L. A fair history of the Web? Examining country balance in the Internet Archive. // Library & Information Science Research. 26(2004) 162-176.

prikupiti. Zato se sve više knjižnica odlučuje za kombinaciju automatskoga i ručnoga načina prikupljanja građe. Pritom pokušavaju što više automatizirati proces prikupljanja dodavanjem posebnih funkcija kod robota, koje filtriraju željenu građu.

Naravno da odluka u većini slučajeva ovisi o mogućnosti knjižnice s obzirom na stručnjake, informacijsko-tehnološku opremu, kao i na pravni kontekst. Sve se više zakona o obveznom primjerku u cijelom svijetu mijenja u smislu obuhvaćanja mrežne građe. U onim zemljama koje još nemaju taj zakon, kriteriji odabira mnogo su stroži jer njihova praksa prikupljanja mrežne građe uključuje i dobivanje autorskih dopuštenja za preuzimanje, smještanje i dugoročnu pohranu elektroničkih zapisa.

U međunarodnoj praksi poznata su tri različita pristupa ili perspektive u definiciji kriterija selekcije: sadržajna, formalna i tehnička.

Sadržajna perspektiva odnosi se na relevantnost sadržaja, što u nacionalnim knjižnicama u većini slučajeva znači sve što je vezano uz određenu zemlju, njezine autore, jezik ili događaje od velikoga nacionalnoga značenja. Ova se perspektiva više prihvaća kod ručnoga ili strogo selektivnoga prikupljanja mrežne građe. Odabir je više ili manje subjektivan. Knjižničari odlučuju o uključivanju građe ako je sadržaj dovoljno reprezentativan i govori o nacionalnoj kulturi, znanosti i društvu, ili ako je autor građe netko od nacionalnih autoriteta. Tu uključujemo i sva tematska prikupljanja mrežne građe. Kod automatskoga prikupljanja nacionalna je domena veoma siguran pokazatelj pripadnosti određenom internetskom prostoru, iako sva prikupljena građa u okviru te domene ne jamči u potpunosti i nacionalni sadržaj.

Formalna perspektiva odnosi se na formalne oblike publikacija na mreži (tipovi publikacija ili "žanrovi"), njihove različite verzije, način pristupa ili prikaz tekstova, dodatne funkcije, koje ovise o računalnom formatu ili razgledniku (*viewer*), izdanja (promjene u tekstovima, obnavljanja baze podataka ...) i slično. I u ovom se slučaju kriteriji selekcije prihvaćaju najviše kod ručnoga prikupljanja. Odlučuje se o formi (naprimjer serijske publikacije, knjige, slike i ostala građa na mreži), o verzijama (naprimjer duplikati u drugim formatima, nova izdanja, duplikati koji su rezultat digitalizacije itd.) ili o porijeklu (original ili rezultat digitalizacije).

Zanimljiv je primjer Nacionalne knjižnice Norveške, u kojoj su za lakše upravljanje mrežnom građom na osnovi Shepherdove i Wattersove²¹³ tipologije formalno odredili tri vrste digitalnih dokumenata:

- tradicionalni - slični su već postojećim tiskanim publikacijama,
- prijelazni - baziraju se na tradicionalnim, iako imaju dodatne funkcije,
- novi - u tiskanim ili drugim oblicima nisu nikad prije postojali.

Tehnička perspektiva kod elektroničke građe tiče se tehničke razine, odnosno platforme strojne opreme, operacijskoga sustava koji omogućava djelovanje aplikacija za pristup, internetskih protokola za prijenos podataka, računalnih formata, načina pristupa, broja datoteka po mrežnom mjestu ili dokumentu, veličine tih datoteka itd. Naprimjer, većina knjižnica prikuplja građu koja je u poznatim datotečnim formatima ili je preuzimaju izravno od samih izdavača, a ne na mreži, jer je taj oblik lakše obraditi, izolirati i omogućiti mu pristup. Kod automatskoga prikupljanja možemo se naprimjer usredotočiti na mrežnu građu koja je objavljena u protokolu HTTP ili FTP, a manje na građu objavljenu u ostalim protokolima, ili možemo ograničiti veličinu datoteka kako ne bismo previše opteretili svoj sustav.

Jednom riječi, kriteriji selekcije ovise o pristupu prikupljanju građe. Pri automatskom prikupljanju građe kriteriji selekcije određuju se s obzirom na funkcionalnost i mogućnost većeg prilagođavanja robota za prikupljanje, dok se pri ručnom građu odabire pojedinačno. U oba slučaja postoje pravila koja se temelje na politici odabira i prikupljanja same knjižnice.

6.2. Međunarodna iskustva pri određivanju kriterija odabira

Svaka nacionalna knjižnica definira svoje kriterije odabira, ovisno o metodi prikupljanja mrežne građe. Najčešće su prisutna tri pristupa: automatski, ručni i kombinirani automatsko-ručni.

²¹³ Shepherd, M. ; Waters, C. Identifying web genre : hitting a moving target. <http://1.1.1.1/390933228/284781912T070130144755.txt.binXMysM0dapplication/pdfXsysM0dhttp://users.cs.dal.ca/~watters/www2004WorkShop/pdfs/4.pdf>. (15.12.2006.)

Automatski pristup

Automatskim pristupom prikuplja se mrežna građa u nacionalnim knjižnicama Norveške, Litve, Finske, Danske i Estonije. Njihov glavni argument je da ne znaju i ne mogu znati koje će informacije, odnosno koji će informacijski izvori biti potrebni budućim generacijama. Usto je ta metoda mnogo jednostavnija i nije skupa. Unatoč tada visokim troškovima pohrane podataka, predviđali su da će se ti troškovi u budućnosti smanjivati. Njihovi kriteriji odabira mogu se sažeti u sljedećim točkama:

- mrežna građa prikuplja se u okviru protokola HTTP ili FTP.
- odabire se mrežna građa u okviru nacionalnih domena (*.se, *.lt, *.no, .fi, *.ee), u okviru protokola HTTP ili FTP prikuplja se građa i u drugim generičkim domenama (*.edu, *.com *.org, *.net, *.web, *.firm, *.store, *.info, *.rec, *.arts, *.nom, *.nu²¹⁴ itd.) s različitim filtriranjima;
- koriste se nacionalnim mrežnim mjestima ili registrima kao početnim popisom URL-a; tako su u danskoj Kraljevskoj knjižnici 2005. godine upotrijebili popis od 607.000 registriranih domena u poduzeću DK-Hostmaster;
- neke periodične informacije na mrežnim stranicama (digitalni časopisi i revije) arhivirane su češće: svaka obnovljena varijanta članka ponovno je arhivirana; dodatci ili promjene također se registriraju;
- većina se koristi protokolom *robots.txt*; iznimka je danska Kraljevska knjižnica, jer su u istraživanju 2003-2004. godine ustanovili da su mnoge značajne stranice zaštićene tom komandom;
- sekundarni prioritet u arhivu imaju publikacije i tekstovi na ostalim internetskim protokolima, izvan HTTP (HTTPS, SNMP itd.), kao što su korisničke skupine (*usenet groups, network news*), diskusijske liste (elektronička pošta preko listservera ili ličnih programa), ili stranice na *gopheru* te ostale informacije o internetu, posebice novi protokoli korištenja; ne uključuju interne baze podataka organizacija, intranete, nedovršene dokumente, internetske igrice i oglase;

²¹⁴ (nu=now koja na švedskom jeziku znači "sada", iako je to i kratica za južnopacifički otok Niue)

- da bi spriječili prevelika opterećenja na posjećenim poslužiteljima i izbjegli kružne klopke (*crawler traps*) prikupljanje robota ograničili su s obzirom na količinu podataka (broj i/ili veličina datoteka); naprimjer u danskoj Kraljevskoj knjižnici ograničili su prikupljanje na 5000 datoteka po domeni;²¹⁵
- u slučaju manjih naroda s jedinstvenim jezikom moguće je ograničiti prikupljanje na osnovi leksičkih svojstava.

Ručni pristup prikupljanju mrežne građe

U slučaju ručnoga i strogo selektivnoga pristupa dobar je primjer praksa australske, novozelandske, njemačke ili kanadske nacionalne knjižnice.

Nacionalna knjižnica Australije prva je u svijetu počela primjenjivati strogo selektivni pristup prikupljanju građe. Tvrdili su da je važnije sačuvati "reprezentativne" stranice australske svjetske mreže koje nude uvidaj u kulturne, povijesne ili znanstvene sadržaje, nego li preuzeti sve bez određenih kriterija. U tu su svrhu osnovali Jedinicu za digitalno arhiviranje, u kojoj je djelovala komisija sastavljena od sedam članova i zaposlenika koji su odlučivali o odabiru građe. Danas je arhiviranje mreže dio aktivnosti Jedinice za katalogizaciju, koja je podređena Odjelu za elektroničku građu.²¹⁶ Budući da je tu aktivnost bilo potrebno usklađivati s devet australskih država, kao i s većim knjižnicama u ostalim australskim državama, izradili su detaljne upute, strategije, priručnike, smjernice itd., koji mogu biti uzor ostalim knjižnicama u svijetu.

U okviru projekta djeluje *Selection Committee on Online Australian Publications*,²¹⁷ koji zajedno s odjelom za elektroničke publikacije ocjenjuje nacionalnu vrijednost građe koja je dostupna

²¹⁵ Za prikupljanje su upotrijebili robot Heritrix, podatke su smjestili na dva poslužitelja s 4GB RAM-a. U tri tjedna prikupili su 138,796,750 datoteka, odnosno 5.3 TB podataka. Nakon detaljne analize ustanovili su da najviše prostora zauzimaju datoteke u grafičkom formatu JPEG, najrasprostranjeniji formati za tekstove bili su HTML i PDF, a jedna mrežna stranica sadržavala je u prosjeku 275 objekata i zauzimala oko 12 MB prostora. Vidi: Andersen, B. The DK-domain: in words and figures. URL: http://netarkivet.dk/publikationer/DFrevy_english.pdf. (15.12.2006.)

²¹⁶ Cathro, W. ; Webb, C. ; Whiting, J. Archiving the web : the PANDORA Archive at the National Library of Australia. A paper presented by Warwick Cathro at the Preserving the present for the future web archiving conference, Copenhagen, 18-19 June 2001. URL: <http://www.nla.gov.au/nla/staffaoer/2001/cathro3.html>. (15.12.2006.)

²¹⁷ Gatenby, P. Digital archiving - developing policy and best practice guidelines at the National Library of Australia. URL: <http://www.icsti.org/icsti/2000workshop/gatenby.html> (15.4.2003.)

preko svjetske mreže. Vode se vrlo strogim kriterijima selekcije, na osnovi kojih su i pripremili detaljne smjernice za odabir građe, koje po potrebi mijenjaju i prilagođavaju novonastaloj građi. Svaki dokument ima više recenzenata. Kad je publikacija lokalizirana, odlučuju o njoj. Ako je odaberu, povežu se s nakladnikom u vezi pridobivanja autorske dozvole. Publikacijama koje su komercijalno dostupne, pristup je ograničen.

Prednosti selektivnoga pristupa su sljedeće:²¹⁸

- svaka jedinica građe koju smještaju u web arhiv ocijenjena je s obzirom na njezinu kvalitetu i tehničke mogućnosti;
- svaka jedinica građe koja se čuva u web arhivu može se u cjelini katalogizirati i zato također postaje dio nacionalne bibliografije; njihovi bibliografski opisi udruženi su s katalogom ostale građe i mogu se jednostavno naći;
- korisnik može pristupiti svim jedinicama građe koje su smještene u web arhivu jer je knjižnica već dobila autorsko pravo za pristup;
- svojstva samostojnih jedinica građe koje se čuvaju u web arhivu su veoma dobro znana voditeljima zbirke i zato je mnogo lakše planirati tehniku za njihovu dugoročnu pohranu;
- mrežne stranice koje su nedostupne robotima prikupljačima identificiraju se u procesu obrade i mogu se prikupiti pomoću drugih metoda i u suradnji s nakladnicima.

U Australiji nemaju sve države jednak zakon o obveznom primjerku i, uz iznimku *Northern Territory*, taj zakon ne uključuje mrežne publikacije. Zato je selektivni pristup prikladniji. U web arhiv uključene su statične i dinamične mrežne publikacije te velik broj publikacija koje više ne postoje na mreži. Ova je strategija bila skupa jer su se prikupljanje i arhiviranje publikacija na mreži odvijali veoma sporo.

Australci imaju detaljno definirane kriterije selekcije koji obuhvaćaju mrežne publikacije i mrežna mjesta o Australiji, od priznatoga australskoga autora ili mrežnu građu australskih autora koja je svojim sadržajima od velike socijalne, političke, kulturne, religiozne, znanstvene ili ekonomske važnosti za Australiju.

²¹⁸ Phillips, M. E. Selective archiving of web resources: a study of acquisition costs at the National Library of Australia. // RLG DigiNews, June 15, 2005, URL: http://www.rlg.org/en/page.php?Page_ID=20666&Printable=1&Article_ID=1749. (15.12.2006.)

Godine 2006. u Nacionalnoj knjižnici Australije ozbiljno su razmišljali o uvođenju automatskoga načina prikupljanja građe.²¹⁹

U **Nacionalnoj knjižnici Kanade** od 2000. godine skupljaju mrežne publikacije. Već su 1994. godine u okviru pilotskog projekta *Electronic Publications Pilot Project* (EPPP) odredili standardne formate, mogućnosti pristupa i načine uređivanja prava na pristup, mogućnosti smještanja, obvezni primjerak itd.

Pri odabiru elektroničkih publikacija na mreži Kanađani vode računa o njihovu kulturnom značenju i o zadatcima nacionalne knjižnice za njihovu pohranu i posredovanje javnosti. Elektroničke publikacije, koje preuzimaju na čuvanje, imaju tri različita statusa s obzirom na njihovu važnost u okviru nacionalne produkcije.²²⁰

- elektroničke publikacije za arhiv (*archived*) – to su najvažnije publikacije i moraju biti stalno dostupne kao dio kanadske kulturne baštine;
- elektroničke publikacije za pristup (*served*) – to su publikacije do kojih je omogućen privremeni pristup, s vremenom su skinute s poslužitelja;
- elektroničke publikacije s poveznicom (*linked*) - one ne čuvaju, već samo omogućavaju URL poveznicu do originalnoga dokumenta.

²¹⁹ Nacionalna knjižnica Australije naručila je prikupljanje *Internet Archive*. Specifikacija eksperimentalnoga prikupljanja bila je sljedeća:

- potpuno preuzimanje po domeni *.au;
- eksperimentalni automatsko prikupljanje po oznakama mrežnog područja (DNS) mrežnih stranica, koje nisu u domeni *.au, iako su povezane s već preuzetim stranicama na australskim IP adresama;
- dubina prikupljanja: cjeloviti sadržaj svih mrežnih mjesta - nije bilo nikakvih ograničenja;
- ograničenje su postavili na maksimalnu veličinu datoteka, koje nisu smjele prijeći 100 MB;
- datoteke, čije je prikupljanje tražilo više od 20 min, bile su isključene;
- poštovalo se je pravilo *robots.txt* (zabrana prikupljanja);
- za početni popis mrežnih mjesta (*seed list*) upotrijebili su URL-e australskih stranica, koje je poduzeće Alexa skupilo u studenom 2004. godine i zatim predalo *Internet Archive*. Ti su URL-i upućivali na 342,296 "hostova". Dodatno je australska Nacionalna knjižnica pripremila popis 530 URL-a koji su sadržavali vladine i državne mrežne stranice i stranice obrazovnih institucija. Alexa je dodala još i popis 230 mrežnih mjesta s više od 50,000 stranica. Tim početnim popisom URL-a željeli su uštedjeti vrijeme prikupljanja velikih komercijalnih mrežnih mjesta i preusmjeriti robot na veće vladine i obrazovne institucije.

²²⁰ National Library Of Canada, October 1998. URL: <http://www.collectionscanada.ca/9/8/index-e.html>. (15.12.2006)

Kriteriji za odabir građe za arhiv (mrežne publikacije i t.i. »elektroničke Canadiane«) jesu sljedeći:

- kanadski izvor, sadržaj ili autor;
- prednost imaju publikacije koje su cjelovite, koherentne i samostojne;
- ne prikupljaju sve verzije jedne publikacije - prednost imaju potpunije publikacije, s dodatnom vrijednošću ili funkcijama;
- ne prikupljaju promocijsku građu, hypertext-linked mjesta, nepotpuna mrežna mjesta za usluge;
- mrežnu građu koja je publicirana izvan Kanade prikupljaju u okviru sadržaja kao što su kanadska glazba i literatura, i to u okviru perspektive utjecaja na društveno-povijesni razvoj kanadskoga društva.

Kod arhiviranja mrežne građe trude se da dobiju publikacije u standardnim formatima, a izbjegavaju konverzije.

U okviru ručnoga pristupa možemo govoriti i o **tematskom pristupu**, iako postoje primjeri kada je građa na određenu temu prikupljena kombinirano, ručno i poluautomatski. Primjer gradnje zbirke u web arhivima na osnovi određene tematike ili događaja nalazimo u američkoj Kongresnoj knjižnici, kao i u Britanskoj knjižnici.

U okviru projekta Kongresne knjižnice MINERVA mrežne stranice i datoteke sa tekstovima koje su dostupne na mreži skupljaju se u cjelini, a zvučni zapisi i videodatoteke ne.

Pri prikupljanju mrežne građe u projektu MINERVA susreću se s određenim teškoćama, kao što su naprimjer:

- teško je odrediti domene osobnih mrežnih stranica;
- teško je odabrati početne stranice kod tematskih prikupljanja mrežnih mjesta;
- teško je namjestiti dubinu odnosno razinu preuzimanja ulaznih i izlaznih poveznica kako ne bi došlo do prikupljanja velikoga broja irelevantnih stranica ili kako se ne bi ispustile stranice koje su vrlo važne;

- teško je prvesti kraju tematsko prikupljanje;
- kvaliteta preuzetih mrežnih stranica mora se provjeravati ručno;
- postoji velika razlika u formatima; svaka datoteka na mrežnoj stranici ima svoj određen format s *MIME type*, naprimjer text/html ili image/jpeg, iako se gotovo svakodnevno pojavljuju novi formati i novije verzije već poznatih formata;
- mnoge datoteke sadrže pogreške i nedosljedne su; ne odgovaraju formatima ili imaju poveznice na datoteke koje više ne postoje; u tom je smislu skoro svaka zahvaćena mrežna stranica zabilježila bar jednu pogrešku;
- vrijeme preuzimanja jedne mrežne stranice ili određene datoteke može trajati čak i više sati;
- pogreške kod preuzimanja mrežnih stranica mogu nastati zbog slabih poveznica ili nedostupnosti poslužitelja;
- teško se preuzimaju baze podataka i dinamične stranice.

Nacionalne knjižnice koje prikupljanje mrežne građe temelje na ručnom principu, prihvaćaju vrlo detaljne specifikacije kriterija odabira mrežne građe koji se odnose na njezin sadržaj i formu. Najiscrpnija pravila ima Nacionalna knjižnica Australije.

Kriteriji selekcije australske Nacionalne knjižnice mogu se sažeti u sljedećim točkama:

- prikuplja se »australika«, dakle sve o Australiji; sve što su australski autori napisali i što ima društvenu, političku, kulturnu, religioznu, znanstvenu ili ekonomsku vrijednost za Australiju; sva djela priznatih australskih autora koji pridonose međunarodnoj znanosti;
- u okviru australike prikupljaju se sljedeći dokumenti: godišnji izvještaji, vladine publikacije, nastavna građa akademskog značenja, građa s izložbi (ako nema tiskanih kataloga), mrežna mjesta etničkih zajednica, sve indeksirane periodične publikacije i članci; vrlo selektivno prikupljaju se mrežne stranice ili mjesta koje su stvorili mladi ili koji su namijenjeni omladini, književna djela, početne stranice osoba i poduzeća te elektronički dokumenti s religioznim sadržajima;
- ako dokument, koji dobiju preko interneta, nema ISBN ili ISSN broja, odabiru ga samo ako odgovara spomenutim kriterijima iz prošle točke, ako ga u budućnosti mogu upotrijebiti kao

uzorak za istraživanje o internetu ili ako dokument može poslužiti kao podatak o tome kako je internet bio upotrijebljen kao sredstvo za publiciranje i širenje informacija;

- ne prikupljaju elektroničke časopise i doktorske disertacije jer o njima u većini slučajeva brinu ostale institucije; jednako postupaju s nacrtima, nezavršenim djelima i oglasima;
- arhiviraju publikacije koje su bitne za istraživanja iako ih već druge institucije arhiviraju; uključuju samo revije s recenzijama;
- u slučaju publikacija koje imaju mnogo unutarnjih i vanjskih poveznica, arhiviraju samo unutarnje poveznice, koje se nalaze na istoj osnovnoj domeni; pregledavaju sve poveznice u dubinu i arhiviraju one koje su bitne za naslov koji arhiviraju;
- ako je publikacija dostupna u različitim medijima, arhiviraju verziju u elektroničkom obliku samo ako on nije reprodukcija, odnosno ako ima dodanu vrijednost; mrežna verzija ima prednost pred verzijom na CD-ROM-u ili disketama.

Većina nacionalnih knjižnica prikuplja tekstove ili slikovnu građu. U slučajevima kada u zemlji nema specijaliziranih institucija za arhiviranje ostalih vrsta građe, kao što su filmovi, zvučne snimke i slično, onda taj zadatak preuzimaju i nacionalne knjižnice (naprimjer Nacionalna knjižnica Estonije). Njemačka nacionalna knjižnica surađuje s akademskim ustanovama i također prikuplja doktorske disertacije, dok Nacionalna knjižnica Australije to prepušta sveučilištima.

Za razliku od Nacionalne knjižnice Australije, u Nacionalnoj knjižnici Novog Zelanda kriteriji selekcije vrlo su pojednostavljeni tako što se elektroničke publikacije definiraju kao različit format publikacija. Zato imaju opće kriterije selekcije za svu građu. Kod gradnje zbirke odlučuju se samo na jedan format, koji ne mora nužno biti elektronički. Digitalni arhiv koji grade čine publikacije od nacionalnog značenja t.j. sva publicirana i npublicirana građa koja obrađuje bilo koji aspekt o Maurima, Novom Zelandu i pacifičkim zemljama te građa iz područja obrazovanja.

Neke nacionalne knjižnice koje imaju vrlo selektivan pristup, naprimjer Nacionalna knjižnica Australije, vode popis prihvatljivih formata. Ostale, naprimjer nizozemska Kraljevska knjižnica, u procesu unosa dokumenata u arhiv pretvara ih u opće prihvaćene formate.

U većini slučajeva prednost se daje mrežnoj građi koja je nastala u elektroničkom obliku. Australci, naprimjer, isključuju reprodukcije tiskanih ili dijela na materijalnom nositelju (skenograme ili elektroničke verzije tiskane građe), osim ako mrežna verzija ne sadrži dodatne značajne funkcije.

Pristup web arhivu uređuje se ovisno o nacionalnoj regulativi i o dogovoru s nakladnicima.

Kombiniran automatsko-ručni pristup

Zbog problema dinamičkih stranica i dubinske mrežne stranice do kojih roboti prikupljači ne stižu ili ih ne uočavaju, vrlo se često kombiniraju kriteriji selekcije za automatsko i ručno prikupljanje. U norveškoj i francuskoj nacionalnoj knjižnici uspjeli su integrirati oba pristupa, dok su u češkoj Nacionalnoj knjižnici oba pristupa odvojena.

Norvežani zahvaćaju publikacije prema sljedećoj metodologiji:

- automatski prikupljaju mrežne stranice na osnovi domene *.no i ostalih predodređenih domena u okviru protokola HTTP, FTP i ostalih;
- zatim izdvajaju mrežne stranice koje odgovaraju nekom od postavljenih kriterija, kao što su različiti događaji ili slično;
- na kraju obrađuju specifične dokumente iz odabranih stranica.

U **Francuskoj nacionalnoj knjižnici** utvrdili su da prikupljanje građe po domenama može stvarati teškoće zbog velikoga broja generičkih domena ili zbog njihove objave u okviru drugih nacionalnih domena. Mnoge firme ili autori koriste se uslugama stranih agencija koje vode registre svojih nacionalnih domena, jer omogućavaju različite igre riječima ili jednostavno nude jeftinije gostovanje. Jezična ograničenja imala bi smisla kad bi se izdvojile domene zemalja francuskoga govornoga područja, kao što su Kanada, Libija i ostale. U mnogim slučajevima potrebna je ručna intervencija.

Svjesni su toga da je mnogo učinkovitije ako se prikupljanje usredotoči na granularnost mrežnih mjesta nego na mrežne stranice. Tako postižu da je cjelina mrežnoga mjesta povezana i da je

detekcija dinamičnih stranica mnogo lakša. Usto je lakše utvrditi je li poslužitelj prestao djelovati. To znači da osim ekstenzivnoga upotrebljavaju i intenzivno prikupljanje.

Izbor opisane metode prikupljanja u Francuskoj nacionalnoj knjižnici argumentiraju problemom svakodnevne promjene mrežnih mjesta. Kao i za današnje istraživače, tako će i za buduće biti važno da po potrebi dobiju iscrpne podatke o svim starijim brojevima elektroničkih časopisa. S ekstenzivnim prikupljanjem koji se odvija nekoliko puta godišnje može se dogoditi da se između prvoga i drugoga prikupljanja mrežna stranica više puta promijeni. Da bi to spriječili, definiraju stranice koje je zbog veće promjenljivosti i relevantnosti (na njih upućuje više poveznica) potrebno češće preuzimati. Usto uvažavaju sljedeća mjerila:

- stranice, čiji tekstovi sadrže manje frekventne riječi, doznaju na specijalizirano područje i one su relevantne za preuzimanje,
- veću težinu daju datotekama s tekstovima nego ostalim oblicima građe, kao što su slike, podatci itd.,
- uzimaju u obzir međusobnu isprepletanost i povezanost mrežnih stranica i stranica s više ulaznih poveznica.

U češkoj Nacionalnoj knjižnici primjenjuju oba pristupa, ovisno o građi. Njihovi su kriteriji sljedeći:

- prikupljanje mreže po najvišim domenama;
- prikupljanje mrežne građe po jezičnom kriteriju - automatski ili ručno;
- prikupljanje mrežnih mjesta s edukativnim, znanstvenim ili umjetničkim sadržajima, koji su od velikog značenja iz nacionalne perspektive;
- prikupljanje mrežnih stranica koje temelje na osnovnim bibliografskim kategorijama;
- elektroničke publikacije koje su slobodno dostupne;
- elektroničke publikacije u općeprihvaćenim formatima (naprimjer: html, xml, jpg i drugi MS Office formati).

6.3. Tumačenje koncepta slovenike

Koncept slovenike sam po sebi označava pojam patriotike, koja obuhvaća iscrpnu zapisanu baštinu jednoga naroda i zato je njezino prikupljanje osnovni i glavni zadatak nacionalnih knjižnica.²²¹ U pravilu su u zbirku patriotike uključena djela koja su izdana u državi, djela, koja se po svome sadržaju odnose na državu i djela autora koji na neki način pripadaju tom narodu. Zato je patriotika u ostalim zemljama dobila ime koje označava nacionalnu pripadnost, naprimjer *croatica*, *polonica*, *hungarica* itd. Koncept patriotike nije u svim narodima identičan i uvjetovan je društveno-povijesnim uvjetima sredine.

Slovenska se patriotika prvi put spominje kao slovenika tridesetih godina prošloga stoljeća. U Zakonu za knjižničarstvo iz 1982. godine.²²² i u novijem Zakonu iz 2001. godine²²³ prikupljanje slovenike jedan je od važnijih zadataka Nacionalne knjižnice. Međutim, kao što utvrđuje Eva Kodrič-Dačić,²²⁴ definicije ni u jednom od tih dvaju zakona²²⁵ nisu pravilne jer ne uzimaju u obzir razlike između nacionalne pripadnosti i državljanstva. Naime, stari i novi zakon u pojam slovenike uključuju literaturu nacionalnih manjina, talijanske, mađarske i romske nacionalne zajednice, dok bi temeljni kriterij slovenike morao biti "slovenski narod, njegov jezik, njegovo kulturno stvaralaštvo i njegovo etničko područje".²²⁶

Slovenika se često miješa s pojmom "nacionalne bibliografije". Nacionalna je bibliografija izbor nacionalne produkcije, patriotike, i mnogo je uži pojam od same slovenike. U Sloveniji su

²²¹ Kodrič-Dačić, E. u djelu Slovenika: strokovne novosti novoga Zakona o knjižničarstvu // Knjižnica, 46, 4(2002), str. 65-85 upotrebljava termin patriotika i vrlo detaljno objašnjava njezin nastanak i razvoj. Termin slovenika poslovenčeni je oblik latinskoga termina "Slovenicae".

²²² Zakon o knjižničarstvu // Uradni list Socialistične Republike Slovenije, 39, 27(1982), str. 1747-1752.

²²³ Zakon o knjižničarstvu // Uradni list Republike Slovenije, 11, 87(2001), str. 8685-8693.

²²⁴ Kodrič-Dačić, E. Slovenika: strokovne novosti novoga Zakona o knjižničarstvu // Knjižnica, 46, 4(2002), str. 65-85.

²²⁵ U novom Zakonu o knjižničarstvu iz 2001. godine piše:

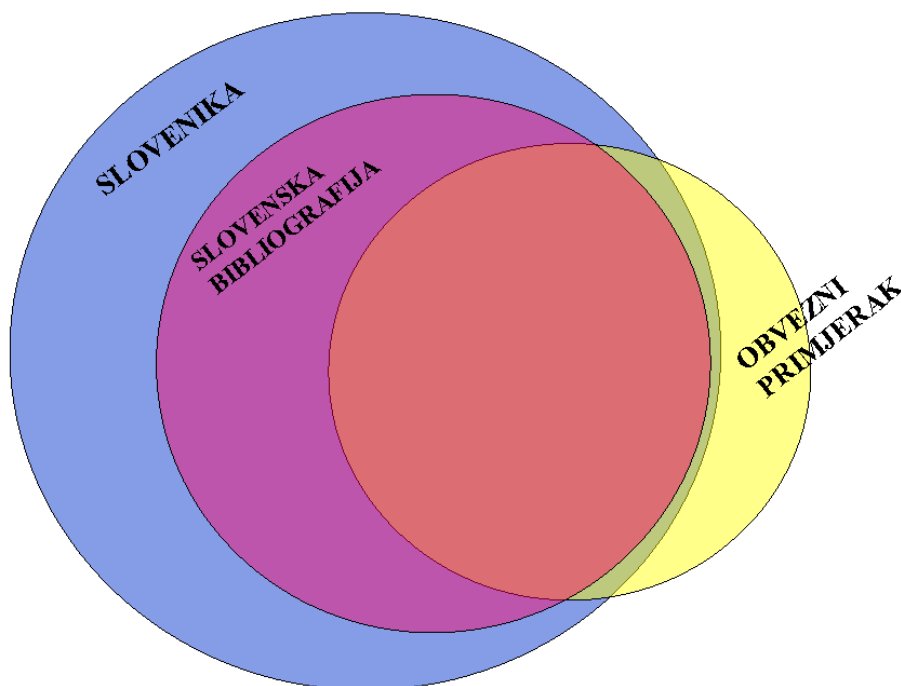
"Nacionalna knjižnica izvaja v okviru knjižnične javne službe (...) naslednje naloge:

- zbira, obdeluje, hrani in posreduje temeljno nacionalno zbirko vsega knjižničnega gradiva v slovenskem jeziku, o Sloveniji in Slovencih, slovenskih avtorjev, slovenskih založb, pripadnikov italijanske in mađarske narodne skupnosti, romske skupnosti in drugih manjšinskih skupnosti v Sloveniji (Slovenika) ter temeljne tuje literature;" članak br. 33 iz Zakona o knjižničarstvu iz Uradni list RS 87/2001, objavljen 8. 11. 2001. - <http://www.uradni-list.si/1/objava.jsp?urlid=200187&stevilka=4446>. (15.12.2006.)

²²⁶ Zakon o knjižničarstvu iz 2001. godine, str. 81.

kriteriji nacionalne bibliografije: slovenski autor, slovenski nakladnik ili slovenski jezik.²²⁷ U novije vrijeme, krajem devedesetih godina, uključuje i elektroničke publikacije na materijalnim nositeljima.

Nacionalnoj knjižnici pri prikupljanju slovenike u velikoj mjeri pomaže Zakon o obveznom primjerku. Međutim, taj zakon nije uvjet, a još manje jedini izvor slovenike, jer je vezan za administrativne granice države i uključuje sve tiskare u zemlji, i one koje tiskaju za strane nakladnike. Često ta literatura ne ulazi u područje slovenike.²²⁸ (vidi sliku br. 9)



Slika br. 9: Odnos građe slovenike, nacionalne bibliografije i obveznoga primjerka

Ako slovenika obuhvaća svu nacionalnu produkciju slovenskoga naroda, onda bi se ti kriteriji morali prenijeti na mrežnu građu. Kod izgradnje zbirke na materijalnim nositeljima u svim je

²²⁷ Kriteriji nacionalne bibliografije temelje se na principu odabira: slovenski autor, jezik ili nakladnik. Obuhvaćaju sljedeće publikacije: elektroničke knjige, serijske publikacije od prve godine izlaženja, članke (selektivno iz časopisa, novina, revija), serijske publikacije koje rjeđe izlaze (kalendari, ljetopisi, zbornici, almanasi) i knjižne zbirke te neknjižnu građu na materijalnom nositelju.

²²⁸ Kodrič-Dačić, E., op.cit. str. 69.

knjižnicama bilo jednostavno odrediti što smije, a što ne smije biti uključeno u nacionalnu zbirku, odnosno što je dio patriotike. U elektroničkom okruženju, s obzirom na količinu građe i na način prikupljanja, nužna su pravila koja bi nam pomogla odabrati samo relevantne i reprezentativne sadržaje pisane baštine. Postavlja se pitanje ima li smisla prenijeti kriterije selekcije tradicionalne građe u elektroničko okruženje, i ako ima, kako je to moguće ostvariti. Pritom se otvaraju nova pitanja koja ukazuju na višedimenzionalnost i višeslojnost elektroničkih publikacija.

6.4. Određivanje kriterija odabira u okviru slovenike

Nakon prihvatanja novoga Zakona o obveznom primjerku,²²⁹ na osnovi njegova 17. članka, NUK je morao najkasnije do listopada 2006. godine pripremiti donošenje pravilnika za prikupljanje publikacija na mreži. Pravilnik o prikupljanju mrežne građe (*Pravilnik o zajemanju spletnih publikacij*²³⁰ - u daljem tekstu Pravilnik) bio je formalni odgovor na zakonski propis i zato se u njemu vrlo općenito spominju kriteriji odabira svih elektroničkih publikacija. U okviru kriterija slovenike određuju se posebni kriteriji koji se odnose na sadržaj, autoritet izdavača ili autora, domene, formate i publikacije, koje su već objavljene na drugim materijalnim nositeljima. Određeni su i tipovi elektroničkih publikacija koje NUK neće prikupljati, među njima su programi pretraživači, računalne igrice, propagandne stranice, stranice koje su rezultat kompilacije podataka, kao što je telefonski imenik i slično, intraneti, baze podataka, diskusijske liste i forumi, liste elektroničke pošte i reprodukcija građe koja već postoji na papiru i kod koje digitalna kopija nema dodatnih funkcija. Jedan od razloga za isključivanje građe je ograničen kapacitet NUK-ova sustava.

Spomenuti Pravilnik sadrži vrlo općenitu smjernicu za bibliotekare NUK-a i ne ulazi u postupke odlučivanja o odabiru elektroničke građe u praksi. Zato će za uspješno izvođenje Pravilnika biti potrebna dodatna specifikacija kriterija odabira u obliku uputa koje bi pomogle pri konkretnom odlučivanju o pojedinim publikacijama na mreži. Pritom su vrlo dragocjena iskustva NUK-a na

²²⁹ Zakon o obveznom izvodu publikacij - ZOIPub. // Uradni list RS 69/06

²³⁰ Donošenje nacrtu Pravilnika bilo je pripremljeno u listopadu 2006. godine. Temeljio se na različitim uzorima u međunaordnoj praksi, između ostalog na hrvatskim Kriterijima odabira obveznog primjerk mrežne građe za obradu i arhiviranje (<http://www.nsk.hr/DigitalLib.aspx?id=83>) te rezultatima projekta *Web Cultural Heritage*.

projektima *Metodologija arhiviranja slovenskih publikacija na mreži* i *Web Cultural Heritage*,²³¹ u okviru kojih je bilo potrebno riješiti problem kriterija selekcije. U specifikacijama kriterija odabira mrežne građe, mora se voditi računa o tri vrste kriterija (sadržajnom, formalnom i tehničkom) u okviru automatskoga i ručnoga pristupa u prikupljanju mrežne građe. Specifični kriteriji odabira ovise i o načina pohrane. Naime, automatski prikupljena građa smješta se u web arhivu, dok se ručno prikupljena građa pojedinačno obrađuje i smješta u repozitorij digitalnog arhiva, a njihova lokalizacija bi bila moguća bibliografskim opisom. U nastavku dajemo kratak pregled tih kriterija.

6.4.1. Kriteriji odabira na osnovi sadržaja

Mrežnu građu prikupljamo na osnovi njezine sadržajne relevantnosti. Pripadnost sloveniki nagovještava slovenski jezik, slovenski autor, slovenski zakladnik ili tematika koja je od važnosti za Sloveniju. Kod različitih pristupa različito se pokušava prikupiti sadržajno relevantna građa.

Automatski pristup prikupljanju mrežne građe:

Kod automatskoga prikupljanja mrežne građe jedan od općih sadržajnih kriterija jest pripadnost mrežne stranice ili mrežnoga mjesta najvišoj nacionalnoj domeni, t.j. *.si. URL-i s ostalim generičkim domenama (*.com, *.net, *.org i sl.) mogu se dodatno definirati uz naknadno fino namještanje s obzirom na mogućnosti robota (leksičke i druge mogućnosti filtriranja).

Relevantnost sadržaja može se odrediti i u okviru mrežnih stranica s nacionalnom domenom koje imaju veću frekvenciju promjenljivosti ili u tekstovima u kojima se koristi veći broj manje frekventnih riječi.

Većina institucija kod namještanja robota poštuje *robots.txt*. Unatoč tomu što su Danci otkrili da je veliki dio relevantne građe skriven iza te komande, nepoštovanje zahtjeva poslužitelja

²³¹ Projekt je financirala Europska Komisija u okviru programa Kultura 2000, glavni koordinatorski projekt bila je češka Nacionalna knjižnica, a kao koordinatori sudjelovali su NUK, Nacionalna knjižnica Estonije i Sveučilišna knjižnica u Bratislavi, Slovačka. Cilj projekta bio je na osnovi međunarodne komparativne analize proučiti i odrediti opće kriterije odabira mrežne građe, koja bi vrijedila za šire europsko okruženje. Osim toga, usporedili smo različite programe za prikupljanje mreže: HTTrack, Heritrix i slovenski WebBird. Projekt se izvodio od 2005. do 2006. godine. Više o projektu u publikaciji *Web Cultural Heritage : Culture 2000 project 25.9.2005-24.9.2006*. Prague : National Library of the Czech Republic, 2006 ili na mrežnoj strani URL: <http://www.webarchiv.cz/culture-2000/>. (15.12.2006.)

moglo bi dovesti do sukoba s nakladnicima. Zato je bolje takvu građu prikupiti ručnim metodama.

Ručni pristup prikupljanju mrežne građe:

Kod ručnoga je pristupa proces prikupljanja subjektivan. Radi lakšega odlučivanja potrebno je odrediti vrste sadržaja koji bi se ručno prikupljali. Prednost bi morala imati građa s trajnom kulturnom, društvenom, znanstvenom ili istraživačkom vrijednošću. Veći dio te građe publiciran je u tradicionalnim oblicima (elektroničke knjige, elektroničke serijske publikacije itd). Među njima su i publikacije koje se nalaze u dubinskoj mreži ili su objavljene na komercijalnoj osnovi, a za koje je potrebno stupiti u kontakt s autorima ili nakladnicima. U slučajevima kad je jedinica elektroničke građe identična tiskanoj građi i nema mogućnost izvođenja dodatnih funkcija, osim te da je virtualno dostupna, postavlja se pitanje je li vrijedna prikupljanja i dugoročne pohrane, s obzirom na to da je njezin životni ciklus mnogo kraći od tiskane verzije.

U svakom slučaju, morali bismo nastojati što više sačuvati nove oblike elektroničkih publikacija koje definira genologija interneta, a koji će vjerojatno odigrati važnu ulogu u budućnosti. Mi knjižničari još uvijek razmišljamo u okvirima tradicionalne građe i te koncepte prenosimo u elektroničku građu. S druge strane, vidjeli smo da se neki tipovi publikacija već mijenjaju i poprimaju nove funkcije i uloge. Takav je slučaj s elektroničkim znanstvenim serijskim publikacijama čije bibliografije i citiranja omogućavaju pristup izvorima informacija na kojima se njezini radovi temelje. Daljnji razvoj tih serijskih publikacija nagovještava nove publikacije iz kojih će budući korisnik imati cjeloviti pregled na razvoj određene ideje, teorije ili koncepta, od njihova nastanka. Sve više ih prate i različite anotacije (*bookmarks*) koje pomažu u vrednovanju kvalitete informacija. Usto postoje i noviji oblici publikacija, kao što su blogovi, tržišne ocjene (*word of mouth* - WOM), česta pitanja (*Frequently asked questions* - FAQ), brze stranice (WIKI) i slično, koje svakako treba prikupiti, posebice one za koje ne znamo u koju ih kategoriju svrstati, jer iz njih mogu nastati novi tipovi publikacija.

Upravo bi tako trebalo sačuvati inovativne primjere publikacija na mreži za koje, zbog eksperimentalne upotrebe novih tehnologija, očekujemo da će imati istraživačku vrijednost u budućnosti.

Prikupljanje bi se moralo usredotočiti na originalnu građu, a ne na površan pregled informacija koji je dostupan na drugim mjestima. Prijevodi dolaze u obzir samo ako odgovaraju navedenim sadržajnim kriterijima.

Za svaki se tip elektroničke publikacije može ograničiti njihovo prikupljanje sadržajnim kriterijima. U nastavku je popis mogućih sadržaja.

Publikacije Vlade Republike Slovenije - Na ovom području može doći do udvostručavanja prikupljene građe između NUK-a i Nacionalnog arhiva Slovenije jer je produkcija Vladinih institucija uvijek bila u ovlasti arhiva. NUK bi morao prikupljati samo one publikacije koje su namijenjene javnosti i koje imaju dugoročnu vrijednost (važeći zakoni, akti, priopćenja za javnost, bilteni, mrežne stranice Vladinih institucija itd.)

Godišnje izvještaje javnih slovenskih organizacija, državnih ustanova, slovenskih sveučilišta i ostalih, NUK prikuplja samo ako nisu dostupni u tiskanom obliku.

Digitalizirana građa - Digitalne reprodukcije građe od nacionalne važnosti, kao što su kodeksi, stara građa, stari zvučni snimci i slično, NUK prikuplja ako nema druge institucije koja bi bila zadužena za tu građu (naprimjer, Nacionalni arhiv, RTV itd.).

Didaktička građa - NUK bi morao prikupljati i građu koju stvaraju obrazovne i akademske institucije, a namijenjena je potrebama obrazovanja.

Doktorske disertacije, magistarski radovi i druga znanstvena, odnosno istraživačka elektronička građa - Znanstveno-istraživačka produkcija Slovenije relevantna je građa i mora biti uključena u digitalni arhiv NUK-a. Sveučilište u Ljubljani namjerava omogućiti

mrežni pristup tim djelima, što time postaje predmet automatskoga prikupljanja na mreži ili ručnoga prikupljanja preko posrednika.

Referati sa stručnih i znanstvenih skupova - Obično se objavljuju na mrežnim stranicama skupova, uz programe. Većinu te građe prikupe roboti. Ručno bi se prikupljanje usredotočilo na pune tekstove referata slovenskih autora, koji bi se odvojeno čuvali i dobili svoj bibliografski opis u katalogu.

Internetske izložbe bit će posebno prikupljene samo ako izložba nema tiskanoga kataloga ili su u internetskoj verziji prikazani dodatni sadržaji kojih nema u tiskanom katalogu.

Elektronički časopisi, elektroničke novine - Svi poznati elektronički časopisi i novine bit će posebno prikupljeni. Prednost imaju znanstveni časopisi s recenzijom.

Mrežne stranice o određenoj temi - Po potrebi će NUK odrediti tematiku u okviru koje će se prikupljati mrežne stranice. Potencijalne teme su predsjednički i stranački izbori te predsjedavanje Slovenije Europskoj Uniji. Moguće su i proizvoljne teme, naprimjer ekološke mrežne stranice ili slovenska kulturna baština na mreži.

Mrežne stranice mladih autora - Te će stranice NUK prikupljati vrlo selektivno i samo ako pokazuju inventivnost kod upotrebe interneta ili prikazuju stajališta mlađe generacije, njihove interese, kulturu i način života.

Beletristika - NUK bi morao prikupiti sva završena književna djela koja su objavljena na mreži. Izbor se može ograničiti samo na ona djela koja imaju veću vrijednost.

Naslovnice organizacija ili pojedinaca - NUK bi morao prikupljati mrežne stranice koje sadrže važne informacije o slovenskim organizacijama (kao što je njihova uloga, projekti, istraživanja, djelatnosti itd.). Osobne bi se naslovnice posebno obradile ako imaju veliku znanstveno-istraživačku ili nacionalnu vrijednost, odnosno kvalitetu koja nije dostupna na drugom mjestu.

Blogovi - Tematski mogu biti vrlo zanimljivi i prikazati različite aspekte društva i kulture kroz specifičnu tematiku.

Baze podataka - Mnoge baze podataka koje su dostupne na mreži dinamičke su i interaktivne te je njihovo automatsko prikupljanje neuspješno. Zato je potreban ručni pristup pri njihovu zahvaćanju.

Osjetljiva građa ili građa koja je pravno sumnjiva također može biti relevantna za različita istraživanja u budućnosti. Prikupljala se samo ako ima političko, povijesno, društveno ili kulturno značenje za Sloveniju. Pristup ovoj građi u arhivu može biti ograničen.

Zemljovid - Elektronički zemljovid ili digitalizirane kopije slovenskih zemljovida predmet su prikupljanja specijaliziranih slovenskih geodetskih, geografskih i ostalih institucija. NUK bi morao prikupljati i pohranjivati elektroničku kartografsku građu samo ako nije dostupna na drugom mjestu, ako su informacije u njoj vjerodostojne i ako je izrađena na osnovi stručnih mjerila kartografske građe.

Notni zapisi bit će prikupljeni samo ako sadržajem pripadaju slovenici i ako su slobodno dostupni na mreži.

Zvučne snimke - NUK će prikupiti zvučne snimke na mreži samo ako sadržajem odgovaraju slovenici i nisu reprodukcija originala objavljenih na materijalnim nositeljima.

Mrežne stranice s vjerskim sadržajima - Prikupljat će se one mrežne stranice koje su odraz stajališta slovenskih zajednica u vezi s vjerskom pripadnošću ili su rezultat dubljih analiza pojedinačnih autora u okviru povijesnih, filozofskih ili teoloških znanosti.

Građa koja dokumentira povijesni razvoj slovenskoga interneta - Tematski bi morale biti odabrane, obrađene i uključene u digitalni arhiv NUK-a one mrežne stranice koje sadrže

zanimljiva i reprezentativna tehnološka, internetska, oblikovna, umjetnička ili bilo kakva inovativna rješenja.

Građa koju NUK ne bi trebalo prikupljati - Tu ulaze nedovršena djela, diskusijske liste, listserveri, računalne igrice, intraneti domaćih organizacija i građa koja nije dostupna široj javnosti, elektronska pošta, programi, dokumentarna građa za koju su odgovorni nacionalni arhivi, neoriginalna djela, programi RTV-a, propagandna građa i građa koja za svoje djelovanje zahtijeva velika sredstva i posebno znanje (poseban softver, informacijska tehnologija itd.).

6.4.2. Kriteriji odabira na osnovi formalnih aspekata mrežne građe

Formalni se aspekti odnose na oblik, funkciju, izgled, izdanje i format mrežne građe. Kod obaju pristupa prikupljanja može se primijeniti kriterije selekcije.

Automatski pristup prikupljanju mrežne građe:

Tip publikacije: Automatskim pristupom nije moguće odrediti formalan tip publikacije. Moguće je filtrirati prikupljanje s formatom datoteka.

Različita izdanja: Prikupljanje mrežne građe trebalo bi se izvesti najmanje četiri puta godišnje. Roboti za prikupljanje mrežne građe sposobni su pronaći mrežne stranice koje se češće mijenjaju. Te bi se stranice trebale češće prikupljati, tako bi se pohranila njihova različita izdanja.

Ručni pristup prikupljanju mrežne građe:

Tip publikacije: Što se tiče određivanja kriterija selekcije s obzirom na tip publikacije, to je prepušteno svakoj pojedinačnoj ustanovi. Najčešće su uključene serijske publikacije, elektroničke knjige, zbornici, istraživački i ostali izvještaji, akademski istraživački radovi, vladini dokumenti i građa koja ima bilo kakvu nacionalnu vrijednost. Isključene su publikacije koje zahtijevaju poseban softver za djelovanje, čije su datoteke veće i teško ih je održavati ili su za njih odgovorni arhivi (e-pošta, baze podataka, intraneti itd.). Publikacije moraju biti cjelovite. Serijske bi se publikacije morale prikupljati periodično.

Različita izdanja: Postavlja se pitanje treba li prikupiti i pohraniti sva izdanja mrežne publikacije, ili pohraniti samo zadnje izdanje, ili pohraniti prvo i zadnje izdanje. Kriteriji odabira na osnovi izdanja morali bi se temeljiti na sadržaju i funkcijama novih izdanja. To znači da bi se morala pohranjivati samo ona izdanja koja imaju veću vrijednost u smislu novih intelektualnih sadržaja ili funkcija. Kod kompilacija i agregatnih baza podataka bilo bi optimalno sačuvati prvo izdanje i svako zadnje nadomjestiti novim.

Problem postojanja tiskanoga i elektroničkoga izdanja. Moguće je da mrežna publikacija već postoji na materijalnom nositelju. Idealno bi bilo pohraniti je zbog omogućavanja mrežnog pristupa. U tom slučaju ne bi bilo potrebno sačuvati sve elektroničke varijante, već samo prvu i zadnju.

6.4.3. Kriteriji odabira na osnovi tehničkih aspekata

Kriteriji odabira koji se temelje na tehničkim aspektima ne ovise o sadržajnim i formalnim kriterijima. Prikupljanje se izvodi na osnovi svojstava informacijsko-tehnološke infrastrukture, funkcionalnosti robota i mogućnosti dugoročne pohrane.

Automatski pristup prikupljanju mrežne građe:

Internetski protokol - Pošto je veći i najrelevantniji dio mrežne građe dostupan preko protokola HTTP ili FTP, dovoljno je da se robot usredotoči na taj dio interneta.

Datotečni formati građe - Prikupljanje bi se moralo usredotočiti na formate, koji su općeprihvaćeni (pdf, doc, html ...); računalne aplikacije s nastavkom *.exe bi bile isključene jer postoji mogućnost da su to nelegalno objavljeni softveri.

Veličina datoteka - Unatoč tome što se svake godine smanjuje cijena prostora za pohranu, troškovi za održavanje veće količine podataka su veći. Zato se mora ograničiti veličina datoteka. Računalne su igrice vezane uz softver i mogu zauzimati mnogo prostora. Obično su dostupne na CD-ROM-ima i kao takve ih nacionalna knjižnica prema zakonu mora dobiti od izdavača. To isto vrijedi i za baze podataka. Većina bitnih dokumenata na mreži ne prelazi 50 MB i zato je zahvaćanje datoteke moguće ograničiti na tu veličinu. Time bi se isključile audio i video snimke koje su u većini slučajeva nelegalne reprodukcije na mreži.

Broj datoteka po mrežnom mjestu - Vrlo malo mrežnih mjesta ima više od 5000 datoteka. Zato bi taj broj mogao biti dodatan filter kod automatskoga prikupljanja. Tako bi se izbjegle i kružne klopke.

Dubina prikupljanja - S prilagođavanjem robota na zahvaćanje do pete razine preuzelo bi se veći dio mreže.

Ručni pristup prikupljanju mrežne građe:

Datotečni formati građe - Kod kriterija odabira format ima vrlo veliku ulogu jer može utjecati na strategije dugoročne pohrane. Mrežna bi se građa morala sačuvati u što više formata, pogotovo ako je njezina relevantnost veća. Međutim, budući da je zbog računalnoga prostora i infrastrukture to rijetko moguće, potrebno je preciznije odrediti kriterij odabira. Pritom može pomoći određivanje formata koji imaju prednost u okviru strategije dugoročne pohrane, za koje će se knjižnica zauzimati (obično su to PDF/A, RTF, TIFF/JPEG2000 itd.). Publikacija bi se morala čuvati u originalnom datotečnom formatu. Ako taj format nije na popisu optimalnih, moralo bi doći do konverzije. Moguće je i da knjižnica izda upute za autore, odnosno da se s autorima ili nakladnicima dogovori o konverzijama.

Materijalni nositelji građe - Ako je publikacija jednako dostupna na mreži i na CD ROM-u, odnosno disketi, prednost će se dati mrežnoj publikaciji jer ima manje tehničkih prepreka za dugoročnu pohranu.

Softver za pristup građi - Elektronička građa mora biti dostupna pomoću opće postojeće opreme i operacijskoga sustava. Elektronička građa koja zahtijeva specijaliziranu i novu opremu ili veću računalnu memoriju bit će prikupljena i pohranjena samo ako je njezina vrijednost veća od troškova njezine pohrane i pristupa.

Osim navedenih specifičnih kriterija odabira, potrebno je odrediti praksu dogovora i uvjeta kod preuzimanja elektroničke građe od zakladnika i autora. Naprimjer, vremensko ograničenje mrežnog pristupa komercijalnih publikacija ne bi smio biti duži od pet godina, dok bi pristup u prostorijama knjižnice morao biti omogućen odmah nakon preuzimanja publikacije bar na jednom računalu ili na određenom broju računala. Baze podataka i publikacije kojima pristup ovisi o posebnom softveru autori ili nakladnici morali bi predati zajedno sa softverom i

odgovarajućim uputama za upotrebu. U slučaju da postoji mogućnost odabira formata publikacije, od njihovih autora ili nakladnika optimalna varijanta.

6.5. Dodatno razmatranje u vezi s kriterijima odabira mrežne građe

Pitanje na koje mi bibliotekari pokušavamo naći odgovor jest: što je zapravo vrijedno sačuvati od svega što je publicirano na svjetskoj mreži? Tu nema kategoričkoga odgovora, odnosno pravi bi odgovor glasio: sve što je moguće, jer ne znamo kako će se dalje razvijati znanost, kakav će biti senzibilitet ljudi u budućnosti, kakav će utjecaj imati informacijska tehnologija na razvoj čovječanstva i što će za naše potomke biti relevantno. Međutim, budući da smo uvijek ograničeni finansijskim sredstvima, brojem stručnjaka, vremenom ili pak informacijskim sustavom, potrebno je odrediti kriterije odabira koji će nam pomoći pri uvođenju sustavnosti i dosljednosti u procesu prikupljanja elektroničke građe.

U određivanju kriterija odabira treba razlikovati dvije razine: kriteriji odabira za prikupljanje građe i kriteriji odabira za obradbu publikacije.

Kod **kriterija odabira za prikupljanje mrežne građe** važno je da što više prikupimo, što znači automatski, ekstenzivno, intenzivno, progresivno, s većom frekvencijom, i dodatno ručno kod publikacija koje se automatskim putem ne mogu prikupiti. Pri automatskom prikupljanju vrlo važnu ulogu imaju funkcionalne mogućnosti robota za prikupljanje mrežne građe u prijenosu kriterija odabira nacionalne produkcije (barem jezik i geografska lokacija).

Kod **kriterija odabira za obradbu mrežne publikacije** možemo primijeniti strože kriterije odabira kako ne bismo previše opteretili sustav. U web arhivu će ionako već biti smještene sve automatski zahvaćene publikacije na mreži i njihovo pretraživanje bit će omogućeno s mrežnim pretraživačima poput Googlea. Zato je potreban dodatni pravilnik o specifičnim kriterijima selekcije ili praktične upute za odabir publikacija na mreži koje će se bibliografski obrađivati. Cilj bibliografske obradbe je učinkovita identifikacija i lokalizacija fonda, a time i njegova veća upotreba. Zato bi obradba morala pratiti potrebe korisnika, odnosno zajednice. Nije potrebno

bibliografski obraditi sve publikacije, koje oblikom sliče tiskanima, jer bi pristup do njih morao biti omogućen preko posebnih sučelja do web arhiva. Obradile bi se komercijalne elektroničke publikacije i publikacije za koje očekujemo da će biti češće tražene u knjižnici, kao što su zakonski propisi ili istraživački radovi. Kompromis između maloga broja raspoloživih knjižničara i potreba korisnika bila bi tematska obradba zanimljivih događaja u slovenskom društvu, kao što su naprimjer politički izbori, predsjedavanje Slovenije Europskoj Uniji i slično. Mrežne stranice i mrežna mjesta skupno bi se obrađivali, a korisnik bi na jednom mjestu mogao naći sve tražene informacije o određenoj temi.

Što se tiče novih žanrova koji već imaju svoj identitet (naslovnica, blogovi, baze podataka, mrežne trgovine) ili ga tek trebaju dobiti (multimedijske enciklopedije, mrežne bibliografije cijelih tekstova, programi za interaktivno učenje na mreži, *hotlists*), postoje dva rješenja: prvo je pričekati s bibliografskim opisom dok ne prihvate i ne traže nove generacije korisnika, a drugo je započeti s bibliografskom obradbom građe najreprezentativnijih i po našoj sadašnjoj ocjeni najvrjednijih žanrova, te tako ubrzati proces njihova uključivanja u tradiciju bibliografske obrade. Pritom može biti od koristi prije spomenuta podjela Shepherd i Watersove.²³²

²³² Shepherd, M. ; Waters, C. Identifying web genre : hitting a moving target. URL: <http://1.1.1.1/390933228/284781912T070130144755.txt.binXMysM0dapplication/pdfXsysM0dhttp://users.cs.dal.ca/~watters/www2004WorkShop/pdfs/4.pdf>. (15.12.2006.)

7. AUTORSKO PRAVO

Jedna je od glavnih obveza svih knjižnica slobodno posredovati i omogućiti pristup pohranjenim informacijama. To je načelo upisano u većini njihovih etičkih kodeksa,²³³ a pravo na informaciju zapisano je u 19. članku Opće deklaracije ljudskih prava.²³⁴ Knjižnice su stoljećima bile arhiv akumuliranoga znanja čovječanstva i u većini slučajeva korisnici su mogli slobodno pristupiti njihovim fondovima. Nacionalnim ili državnim knjižnicama već je u 16. stoljeću zakon obveznoga primjerka dodjeljivalo posebno pravo na prikupljanje i pohranu svih publikacija koje su bile tiskane u određenom području ili kraljevini. Nacionalne knjižnice ne samo da su do danas održale tu tradiciju, već je sva građa i dalje dostupna korisnicima.

S razvojem informacijske tehnologije pisana je građa prešla iz analognih medija na elektronički medij. Elektronička je građa rezultat evolucije tiskarstva. Knjižnice su tako nastavljale prikupljati i posuđivati građu na novijim medijima jednako kao i tradicionalnu građu. Problem je u tome što je pojava nove građe s novim svojstvima utjecala na uvođenje promjena u zakonodavstvu o autorskoj zaštiti. Tako je došlo do kolizije dvaju područja koja su bila definirana zakonom: s jedne strane zakonski regulirana obveza knjižnica²³⁵ da omogućavaju svakom čovjeku da ostvari svoje pravo na pristup informaciji, a sa druge pravo autora i nakladnika da zaštite svoja djela od slobodnoga korištenja.²³⁶

Zastarjeli zakoni o obveznom primjerku u većini europskih država nisu uključivali elektroničku građu. Budući da se zakonodavstvo sporije mijenja od informacijske tehnologije, u takvoj

²³³ Etički kodeks slovenskih knjižničara: <http://www.nuk.uni-lj.si/nuk3.asp?id=37119387> : U 8. članku navodi: "Knjižničar se zavzema za prost pretok gradiv in informacij, vendar ne odgovarja za posledice, ki izhajajo iz njihove uporabe, a 9. članu kodeksa piše: Knjižničar nasprotuje vsem poskusom uvajanja cenzure in drugih strokovno neutemeljenih omejitev pri pridobivanju in posredovanju informacij. "

²³⁴ Universal declaration of human rights, 19. članak glasi: "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers". URL: <http://www.lonweb.org/onu/hr-ing-slo.htm> (15.12.2006.)

²³⁵ Zakoni o knjižničarstvu reguliraju obveze knjižnice, među njima je i omogućavanje pristupa pohranjenim informacijama korisnicima.

²³⁶ Jakac-Bizjak, V. Problematika avtorskih pravic v zvezi z zbiranjem in hranjenjem elektronskih publikacij ter z njihovim dostopnostjo za javnost. // Kavčič-Čolić, A. ... [et al.]. Metodologija zbiranja in arhiviranja slovenskih elektronskih publikacij na medmrežju / [nositeljica projekta Eva Kodrič-Dačić ; projektna skupina Alenka Kavčič-Čolić (rukovodilac projekta) ... et al.]. Ljubljana : Narodna in univerzitetna knjižnica, 2004. (Ciljni raziskovalni programi).

pravnoj praznini nacionalne su knjižnice smjele prikupljati, s iznimkama, i elektroničku građu na materijalnim medijima. Europske su smjernice na tom području štitile autore i vlasnike autorskih prava. Zato su nacionalne knjižnice nastojale da se zakonodavstvo u okviru obveznoga primjerka prilagodi novonastalim situacijama. Taj trend još nije završen u svim europskim zemljama.

U okviru europskog programa CoBRA+ osnovana je zajednička komisija koja je predstavljala Konferenciju direktora europskih nacionalnih knjižnica (*Conference of Directors of European National Libraries* - CENL) i Udruženje europskih nakladnika (*Federation of European Publishers* - FEP). Glavni zadatak te komisije bio je da nađe način sudjelovanja koji bi bio prihvatljiv za obje strane, kako za nacionalne knjižnice tako za nakladnike. Rezultat tih pregovora bio je "Kodeks prakse dobrovoljnog depozita" (*Code of practice for the voluntary deposit of electronic publications*) u daljnjem tekstu Kodeks. Cilj Kodeksa bio je predložiti osnovu u rješavanju problematike obveznoga primjerka publikacija. U Kodeksu su bili definirani svi elementi koje bi trebalo uzeti u obzir pri sklapanju ugovora između knjižnica i nakladnika.²³⁷ Međutim, Kodeks se odnosio na elektroničke publikacije pojedinačnih nakladnika i nije rješavao problem automatskoga preuzimanja mreže.

U Velikoj Britaniji, Njemačkoj i Nizozemskoj Kodeks je osnova za pregovaranje s nakladnicima. Među prvim europskim državama koje su svu elektroničku građu (i mrežnu građu) uključile u svoj zakon o obveznom primjerku je Norveška (1990. godine), a zatim dolazi Hrvatska,²³⁸ koja je 1997. godine donijela Zakona o knjižnicama, u kojem je poseban članak posvećen obveznom primjerku. Francuska, Danska i Australija su u zakon o obveznom primjerku postupno uključivale elektroničke publikacije, i to prvo na materijalnim medijima, a poslije i na mreži. Njemačka je mrežnu građu uključila u zakon o obveznom primjerku 2002. godine, a Velika Britanija 2003. godine²³⁹

²³⁷ Elektronske publikacije. Kodeks prakse prostovoljnega depozita / [tekst priredila i dopunila Vilenka Jakac-Bizjak]. Ljubljana : Narodna in univerzitetna knjižnica, 2001 (Ljubljana : Pleško). 15 str.

²³⁸ Članak 39. Obvezni primjerak Zakona o knjižnicama. Narodne novine 105(1997); 5(1998); 104(2000)

²³⁹ Jakac-Bizjak, V. Problematika avtorskih pravic v zvezi z zbiranjem in hranjenjem elektronskih publikacij ter z njihovim dostopnostjo za javnost. // Kavčič-Čolić, A. ... [et al.]. Metodologija zbiranja in arhiviranja slovenskih elektronskih publikacij na medmrežju / [nositeljica projekta Eva Kodrič-Dačić ; projektna skupina Alenka Kavčič-Čolić (rukovodilac projekta) ... et al.]. Ljubljana : Narodna in univerzitetna knjižnica, 2004. (Ciljni raziskovalni programi)

U većini europskih država zakon o obveznom primjerku daje knjižnicama pravo samo da prikupljaju i pohranjuju elektroničku građu na materijalnim nositeljima. Ne daje im pravo da prikupljaju građu na mreži, a još manje da tu građu posuđuju korisnicima ili je mijenjaju zbog primjene različitih strategija dugoročne pohrane.

Za prikupljanje elektroničke građe na mreži pravna regulativa razlikuje tri različite vrste prava: pravo na preslikavanje iz mreže, pravo na omogućavanje javnog pristupa građi i pravo na intervencije, koje zahtijevaju različite strategije dugoročne pohrane (migracije i konverzije).

U većini država zakon o obveznom primjerku daje nacionalnim knjižnicama pravo na prikupljanje građe. Ostale su djelatnosti autorski zaštićene i sve ovisi o lokalnim interpretacijama zakona. U nastavku su razmotrena različita pravna stajališta.²⁴⁰

7.1. Pravo na prikupljanje mrežne građe

Autorsko pravo (*copyright*) različito se tumači u SAD-u i Europi. Dok američki zakon definira autorsku zaštitu kao "ograničen monopol koji je stvoren s ciljem da ljudi dobe financijsku potporu za stvaranje autorski zaštićenih djela, da stvaraju literarna, umjetnička djela ...",²⁴¹ u Europi autorska zaštita znači "pravo svakoga autora da kontrolira reprodukciju proizvoda svog razuma [...] i] sagledana je [...] kao produžetak autorske ličnosti".^{242, 243} Obje definicije tretiraju različite perspektive jednakog koncepta.

Osnove današnje autorske zaštite određene su na međunarodnoj konferenciji u Bernu 1886. godine. Zadnja revizija te konvencije bila je u Parizu 1997. godine. Najvažniji članak koji se odnosi na autorsku zaštitu (*copyright*) je 9.(2), u kojem se navodi da "nacionalna zakonodavstva

²⁴⁰ Kavčič-Čolić, A. Archiving the web - some legal aspects. Referat predstavljen na 68th IFLA Council and General Conference, August 18-24, 2002. Također objavljeno u: *Library Review*, 52, 5(2003), str. 203-208.

²⁴¹ "the limited monopoly created for the purpose of providing people with a financial incentive to create copyrightable materials, to create works of literature, art,..."

²⁴² "the right of every author to control the reproduction of the products of his or her own brain [...] and] is seen [...] as an extension of the author's personality"

²⁴³ Uspoređenje je uzeto iz Strong, W. S.: Copyright in the new world of electronic publishing. Presented at the workshop Electronic Publishing Issues II at the Association of American University Presses (AAUP) Annual Meeting, June 17, 1994, Washington, D.C. URL: <http://www.press.umich.edu/jep/works/strong.copyright.html>

moгу dozvoliti reprodukciju autorski zaštićenih djela u 'posebnim primjerima pod uvjetom da takva reprodukcija ne ulazi u konflikt sa normalnom eksploatacijom dijela i znatno ne ošteti legitimne interese autora'.²⁴⁴ Godine 1967. osnovana je Svjetska organizacija za intelektualno vlasništvo (*World Intellectual Property Organisation* - WIPO), koja je preuzela odgovornost za administriranje mnogih međunarodnih konvencija i ugovora o intelektualnom vlasništvu i autorskom pravu. Među njima je bila i Bernska konvencija. Najveća konferencija WIPO održala se u prosincu 1996. godine u Ženevi. Usredotočila se na moguće promjene u Bernskoj konvenciji. Između ostalog predloženo je uključivanje prava korisnika da bez dopuštenja pregledavaju građu na računalnim zaslonima ("*browsing, concerning viewing of screened material without permission*").²⁴⁵ Kao što se moglo očekivati, nakladnički lobiji bili su moćniji i to pravo nije bilo uključeno u Direktivu 96/9/EC²⁴⁶ Europske unije o legalnoj zaštiti baza podataka, a još manje u Direktivi 2001/29/EC²⁴⁷ o usklađivanju određenih aspekata autorskoga i srodnih prava. U zadnjoj se navode sljedeća prava autora, citirano:

- "[...] pravo da autorizira ili zabrani izravnu ili posrednu, privremenu ili trajnu reprodukciju na svim medijima i oblicima, u cjelini ili dijelom; (2. članak)
- [...] pravo da autorizira ili zabrani bilo kakvo komuniciranje svojih djela javnosti, putem žice ili bežično, uključujući i mogućnost da svoja djela čine dostupna javnosti tako da članovi publike mogu pristupati tim djelima s mjesta i u vrijeme koje sami odaberu; (3. članak, 1. alineja)

²⁴⁴ "the national legislatures may authorise the reproduction of copyright works in 'certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interest of the author". Wall, Raymond A.: Copyright made easier. - 2nd ed.. - London : Aslib, 1998. - str. 338

²⁴⁵ Wall, R. A. Copyright made easier. 2nd ed. London : Aslib, 1998. str.339.

²⁴⁶ Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. In Official Journal L 077, 27/03/1996 str. 0020-0028. URL: <http://europa.eu.int/eur-lex/lex/Notice.do?val=344298:cs&lang=en&list=391654:cs,391398:cs,276720:cs,276918:cs,284595:cs,266982:cs,272215:cs,236779:cs,344298:cs,415100:cs.&pos=9&page=2&nbl=27&pgs=10&hwords=directive~databases~&checkte=checkbox&visu=#texte> (15.12.2006.) u slovenskom: <http://eur-lex.europa.eu/LexUriServ/site/sl/dd/13/15/31996L0009SL.pdf> (15.12.2006)

²⁴⁷ Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society. In: Official Journal L 167, 22/06/2001, p.0010-0019, URL: http://eur-lex.europa.eu/LexUriServ/site/en/oj/2001/l_167/l_16720010622en00100019.pdf (15.12.2006.)

- [...] *ekskluzivno pravo da autoriziraju ili zabrane svaki oblik distribucije u javnosti putem prodaje ili na drugi način [...]*.²⁴⁸ (3. članak, 2. alineja)

Peti članak te iste Direktive navodi iznimke u slučaju "javnog interesa sa ciljem obrazovanja i u pedagoške svrhe" ("*...in the public interes for the purpose of education and teaching*") i dodatno "za privatnu upotrebu i za namjene koje nisu izravno ili posredno vezane za komercijalne interese i pod uvjetom da vlasnici autorskih prava dobiju poštenu kompenzaciju" ("*... for private use and for ends that are neither directly nor indirectly commercial, on conditions tha the rightholders receive fair compensation*"). Reprodukcijska je ograničena na dijelove dokumenata.

Zakonodavstvo Europske Unije autorski štiti sljedeća djela: računalne programe, sučelja, baze podataka i sva autorska djela bez obzira na medij na kojem su ostvarena. Mrežne stranice se u ovim direktivama ne spominju. Međutim, po analogiji se mogu svrstati u autorska djela i na njima se može primijeniti autorska zaštita. U Velikoj Britaniji Agencija za autorske licence (*The Copyright Licencing Agency*) navodi: "Svjetska mreža je predmet autorske zaštite i zato su mrežne stranice literarna djela".²⁴⁹ Svako prikupljanje ili arhiviranje te građe bez dopuštenja autora ili vlasnika autorskih prava je nelegalno. Iznimke su osobni podatci, vladine publikacije koje pripadaju javnoj domeni i publikacije u kojima je jasno navedeno da je njihova reprodukcija dopuštena. Autorska djela su zaštićena na rok od sedamdeset godina poslije smrti autora.

Budući da mnoge računalne aplikacije za pregledavanja građe zahtijevaju njihovo preslikavanje na zaslon računala, Direktiva 96/9/EC dopušta privremenu reprodukciju u smislu kaširanja (*caching*), gostovanja ili pregledavanja. Aplikacije sa licencom su predmet ugovora sa vlasnicima autorskih prava.

²⁴⁸ "[...] *right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part*; (2. članak)

- [...] *right to authorise or prohibit any communication to the public of their works, by wire or wireless means, including the making available to the public of their works in such a way that members of the public may access them from a place and at a time individually chosen by them*; (3. članak, 1. alineja)

- [...] *the exclusive right to authorise or prohibit any form of distribution to the public by sale or otherwise [...]*. (3. članak, 2. alineja)

²⁴⁹"*The World Wide Web is subject to copyright, and Web pages are themselves literary works*". URL: <http://www.cla.co.uk/copyrightvillage/internet.html> (15.12.2006.)

7.2. Pravo javnosti na pristup arhiviranim mrežnim stranicama

Osim Univerzalne deklaracije ljudskih prava i etičnih kodeksa knjižnica, različite agencije u okviru IFLA-e i drugih međunarodnih agencija pokušavaju dokazati da je pristup informacijama neosporno pravo svakog pojedinca i komercijalni interesi ne bi smjeli pri tome interferirati.

IFLA-in Odbor za slobodan pristup informacijama i slobodu izraza (*IFLA Committee on Free Access to Information and Freedom of Expression - FAIFE*)²⁵⁰ navodi da knjižnice skupljaju, pohranjuju i omogućavaju pristup svim pohranjenim informacijama.²⁵¹ ECUP (*The European Copyright User Platform*) također je prihvatio dokument sa stajalištem (*Position Paper*) u kojem pokušava opravdati važeće zakonite upotrebe autorskih djela sa strane pojedinaca i knjižnica u elektroničkom okruženju.²⁵²

Kao što smo vidjeli, treći članak Direktive 2001/29/EC zabranjuje sve oblike komuniciranja s publikom ili omogućavanja pristupa publici do autorski zaštićenih djela na bilo kakvom mediju bez dopuštenja autora. U svome 5. članku iznimno dopušta specifične oblike reprodukcija knjižnicama, obrazovnim institucijama, muzejima ili arhivima, koji nemaju za cilj posredni ili neposredni komercijalni interes. Ta je iznimka u skladu s anglosaksonskom doktrinom *fair dealing* (*fair use* u SAD-u), čiji je cilj postići ravnotežu između koristi vlasnika autorskih prava i korisnika.²⁵³

²⁵⁰ URL: <http://www.faiife.dk/> (15.12.2006.)

²⁵¹ - "Libraries provide access to information, ideas and work of imagination. They serve as gatewayst to knowledge, thought and culture."

- "Libraries have a responsibility both to guarantee and to facilitate access to expressions of knowledge and intellectual activity. To this end, libraries shall acquire, preserve and make available the widest variety of materials, reflecting the plurality and diversity of society."

²⁵² ECUP: Position on user rights in electronic publications. // *The Liber Quarterly*, 7(1997), str. 386-395 ; Harper, G. Will we need fair use in the twenty-first century? 2001. URL: http://www.utsystem.edu/agc/intellectualproperty/fair_use.htm (15.12.2006.), str. 386.

²⁵³ Američki *fair use* zamišljen je kao "ograničavanje količine građe koju je dopušteno kopirati ili upotrijebiti pošteno (*fair*) tako da ne ošteti interese vlasnika autorskih prava". Vidi : Pedley, P. *Copyright for library and information service*. London : Aslib, 1998.

Britanski *fair dealing* pokušava balansirati s jedne strane prava autora da imaju ekskluzivnu kontrolu njihovih djela, a s druge prava korisnika da slobodno pristupaju sadržaju tih djela. Vidi: Harper, Georgia: *Will we need fair use in the twenty-first century?*, 2001. - URL: http://www.utsystem.edu/agc/intellectualproperty/fair_use.htm. (15.12.2006.)

7.3. Pravo na dugoročnu pohranu mrežne građe

O strategijama i metodama dugoročne pohrane građe bit će govora u drugim poglavljima. Činjenica je da se informacijske tehnologije vrlo brzo mijenjaju, a materijali koji nose digitalne zapise sve brže zastarijevaju. Zato je potrebna redovita migracija podataka, kao i njihova konverzija u novije verzije softverskih aplikacija. U prvom slučaju radi se o reprodukciji, a u drugom o intervenciji u samu logičku strukturu dokumenta. Nažalost europske se direktive ne bave tom problematikom i zato je ona prepuštena nacionalnim knjižnicama i nacionalnim regulativama. Većina nacionalnih knjižnica već izvodi strategije dugoročne pohrane i ne obazire se na zakone o autorskim pravima. S druge strane zakonodavne vlasti zatvaraju oči jer su njihovi postupci u interesu pohrane nacionalne baštine.

7.4. Zakonska rješenja arhiviranja mrežne građe u Sloveniji

Prije prihvaćanja novoga Zakona o obveznom primjerku u Sloveniji je važio Zakon o autorskom i sličnim pravima,²⁵⁴ koji je bio sasvim usklađen s europskom Direktivom. Dotad je svako prikupljanje mrežne građe bilo nelegalno, a kad smo za eksperimentalne svrhe zatražili od javnih institucija da nam dopuste testiranje robota, dopustili su nam to pod uvjetom da u slučaju uvođenja prakse arhiviranja mrežne građe potpišemo s njima poseban ugovor.

Zakon o obveznom primjerku²⁵⁵ pojednostavio je proces arhiviranja mrežne građe i dodijelio je NUK-u sve ovlasti glavne depozitarne organizacije u Sloveniji. Druga depozitarna organizacija je Sveučilišna knjižnica u Mariboru.

Elektroničke publikacije definirane su u 2. članku tog zakona kao "elektroničke knjige, elektronički časopisi i novine, mrežne stranice i slično, koji su objavljeni na materijalnim nositeljima (naprimjer na magnetnim vrpcama, kasetama, disketama, CD-ROM-ima i slično), ili su dostupni preko računalnih mreža ili na svjetskoj mreži."

²⁵⁴ Zakon o avtorski in sorodnih pravicah - uradno prečišćeno besedilo /ZASP-UPB1/ // Uradni list RS, 94(2004). URL: http://zakonodaja.gov.si/rpsi/r08/predpis_ZAKO4148.html (15.12.2006.)

²⁵⁵ Zakon o obveznem izvodu publikacij (ZOIPub) // Uradni list RS.69(2006), str. 7230.

Obveznici davanja obveznoga primjerka su slovenski nakladnici ili distributeri, ako je i kada je publikacija namijenjena za distribuciju u Republici Sloveniji. Obvezni primjerak obuhvaća sve publikacije koje su napisane u slovenskom jeziku ili su prilagođene za slovenski jezik (prijevodi multimedijских sadržaja) na svim materijalnim nositeljima. Prema istom, 4. članku u obvezne primjerke spadaju i publikacije na materijalnim nositeljima koji će se u budućnosti pojaviti kao posljedica njihova razvoja.

Zakon u 6. članku predviđa obvezni primjerak jedne kopije elektroničkih publikacija na računalnoj mreži ili svjetskoj mreži. Elektroničke publikacije moraju biti predložene bez elektronske zaštite pristupa ili nakladnici i ostale odgovorne institucije moraju osigurati depozitarnoj organizaciji lozinku za pristup za njihovo preuzimanje (7. čl. 1. al.). Sedmi članak omogućava NUK-u da daje javni pristup arhiviranoj građi. Međutim, da bi se zaštitili komercijalni interesi nakladnika kod komercijalne građe, isti članak daje mogućnost dogovora s vlasnicima autorskih prava tako da se nađe rješenje u kojem će NUK dopustiti pristup javnosti do građe u prostorijama knjižnice za studijske i istraživačke namjene. Moguće su i varijante vremenskoga ograničenja pristupa.

Članak 17. određuje prikupljanje obaveznoga primjerka elektroničke građe na mreži. U tom smislu predviđa Pravilnik koji će NUK pripremiti i u kojem će specificirati frekventnost preuzimanja mrežne građe i građe zaštićene lozinkama, kao i definirati prikupljanje različitih verzija građe.

U 19. i 20. članku zakon dopušta NUK-u da zbog propadanja ili zastarjelosti tehnologije primijeni različite strategije u svrhu pohrane sadržaja građe, kao što su migracije nositelja i konverzije.

Na pripremu donošenja nacrtu novoga Zakona o obveznom primjerku utjecao je prijedlog Zakona koji je pripremila Ministarstvo za kulturu u suradnji za NUK-om. U tom smo prijedlogu pokušali riješiti svu problematiku arhiviranja elektroničkih publikacija. Većina tih prijedloga bila je prihvaćena. Pritom je dosta utjecao i izvještaj projekta *Metodologije arhiviranja slovenskih publikacija na mreži*, u kojem su obrazloženi različiti aspekti prikupljanja građe.

U svakom slučaju, elektronička građa zbog specifičnih svojstava zahtijeva dodatnu pažnju i u smislu dobivanja autorskih prava i njihove zaštite. Ako je publikacija namijenjena tržištu, onda je potreban ugovor između nakladnika ili autora i knjižnica. Većina knjižnica postiže dogovor za omogućavanje pristupa komercijalnoj građi u lokacijama knjižnice. Kod javno dostupne građe, koja nije namijenjena prodaji, morala bi se uvažavati etična načela tako da pohrana i pristup građi ne štete nikomu. U *Internet Archive* naprimjer postoji pravilo da prema želji pojedinaca ili poduzeća cenzuriraju pristup spornim stranicama u javnosti.

8. BIBLIOGRAFSKA OBRADBA MREŽNE I ELEKTRONIČKE GRAĐE

Delsey²⁵⁶ je 2001. godine preispitao konvencionalne paradigme za opis elektroničke građe. Problem je taj da se nakon dugogodišnje prakse u opisivanju tradicionalne građe i djelovanja u analognom prostoru knjižničari odjednom susreću s građom koja djeluje u sasvim različitom okruženju, ima druga svojstva i ponaša se različito od materijalne građe. Po Delseyevim riječima, u digitalnom se okruženju postavljaju nova pitanja na koja će knjižničari morati naći odgovor:

1. Na čemu temeljimo opis digitalnog dokumenta? Informacijske su tehnologije sposobne za sve veće personalizacije sadržaja iz perspektive prilagođavanja pristupa građi različitim sučeljima, prikazivanje različitih multimedijских sadržaja, omogućavanje funkcionalnosti, ovisno o podršci softverskih programa, i davanje različite jezične mogućnosti pregledavanja sadržaja. Moguć odgovor na ovo pitanje Delsey nalazi u modelu IFLA FRBR,²⁵⁷ tj. u prijenosu fokusa opisa na entitet *djela* i *manifestacije*, dok bi za entitet *izraz* opisali attribute djela, kao što su jezik ili dodatna specifična svojstva. Opis građe time bi bio fleksibilniji i omogućio bi opis svih atributa dokumenta.
2. Kako definirati granice digitalnoga dokumenta? Granice između bibliografskoga opisa, bibliografske reference i kataloga, pa i između dvaju ili više dokumenata, za razliku od tradicionalne građe, u digitalnom se okruženju gube. Time se ruši cijela struktura na kojoj se temelji konvencionalan bibliografski opis. U digitalnom okruženju opis više ne možemo temeljiti na materijalnim svojstvima, a razlika između materijalnoga i intelektualnoga nestaje.
3. Kako možemo najučinkovitije prikazati odnose među dokumentima? U digitalnom okruženju gube se hijerarhijski odnosi. Mnogi odnosi, koji su upotrijebljeni za definiciju materijalnih identiteta između različitih sadržaja, sada više nisu potrebni. Osim toga, mnogi vanjski odnosi

²⁵⁶ Delsey, T. Preispitivanje konvencionalnih paradigmi za opis dokumenata = *Reassessing conventional paradigms for document description* / prevela Mirna Willer // Vjesnik bibliotekara Hrvatske, 46, 1-2(2003), str. 32-43. Original u engleskom dostupan na URL: http://w3.uniroma1.it/ssab/er/relazioni/delsey_eng.pdf, objavljen u zborniku međunarodne konferencije: Le risorse elettroniche : definizione, selezione e catalogazione : atti del convegno internazionale, Roma, 26-28 novembre 2001 / a cura di Mauro Guerrini con la collaborazione di Stefano Gambari e Lucia Sardo. Milano : Editrice Bibliografica, 2002. (Bibliografia e biblioteconomia)

²⁵⁷ Funcional requirements for bibliographic records je model koji se temelji na četiri entiteta: dijelo, izraz, manifestacija, izdanje. Vidi: IFLA Study Group on the Functional Requirements for Bibliographic Records : Funcional requirements for bibliographic records : final report. München : Saur, 1998.

izraženi po FRBR kao "*odnosi izdanje-prema-izdanju između konvencionalnih dokumenata možda bi se trebali odražavati kao unutarnji atributi digitalnog dokumenta*"(str. 39)

4. Jesu li tradicionalni elementi opisa građe prihvatljivi za digitalno okruženje ili su potrebni dodatni podatci? Delsey predlaže promjenu fokusa s otkrivanja digitalne građe prema upravljanju digitalnom građom. Dok je katalog zajedno sa signaturom, koja je pomogla pri lokalizaciji građe, bio posrednik između korisnika i građom, u digitalnom okruženju korisnik neposredno pristupa digitalnoj građi.

Dodatno se u digitalnom okruženju pojavljuju pitanja čuvanja trajnosti veze između opisa građe i samog dokumenta. Postavlja se i pitanje kako ponuditi korisnicima kontekstualne informacije, koje će im ocijeniti relevantnost mrežne građe te kako olakšati sučeljavanje opisa dokumenata koji potječu iz različitih izvora.

Mrežna građa dobiva na važnosti i stručnjaci se sve više bave problemom kako uključiti te publikacije u nacionalne bibliografije. U okviru IFLA-e vodi se rasprava i pokušava se naći način da se ta građa također uključi u nacionalne bibliografije.²⁵⁸ Međutim, njezina neopipljivost otežava bibliografski opis. Postavlja se pitanje što je knjiga na mreži? Kakva je razlika između članka, referata na konferenciji i knjige na mreži? Hoće li se bibliografski opis temeljiti na veličini bajtova ili na broju datoteka? Kakva je razlika među višedijelnoga djela, elektroničke knjige koje se stalno obnavljaju i periodične publikacije, te između naslovnice i serijske publikacije itd.

Christian Lupovici²⁵⁹ na osnovi iskustava u francuskoj Nacionalnoj knjižnici također analizira problematiku bibliografskoga opisa mrežne građe. Metapodatci bibliografskoga opisa tradicionalno su se koristili za opis materijalne građe kako bi olakšali njihovu identifikaciju i traženje. Zato je tradicionalna katalogizacija ručna i prilagođena je materijalnim objektima. Digitalni objekti već u sebi sadrže dokumentaciju (metapodatci o tehničkim svojstvima ili

²⁵⁸ Jacobsen, G. Universal Bibliographic Control and International MARC Core Programme (UBCIM) : International Conference on National Bibliographic Services Copenhagen, 25-27 November 1977: COVERAGE. // ICNBS, Copenhagen, 25-27 November 1998. URL: <http://www.ifla.org/VI/3/icnbs/fina.htm> (15.12.2006.)

²⁵⁹ Lupovici, C. Web crawling : The Bibliotheque nationale de France experience. // World Library and Information Congress: 71th IFLA General Conference and Council "Libraries - A voyage of discovery", August 14th-18th 2005, Oslo, Norway.

autorskim pravima) koja olakšava njihovo upravljanje u sustavu za arhiviranje. Osim toga, digitalni objekti koji temelje na tekstu mogu se indeksirati ili sortirati bez ikakve ljudske intervencije. Nedostatak u metapodacima može se nadoknaditi sadržajem ili kontekstom u internetu. Lupovici dalje navodi da time ne misli da deskriptivni metapodatci ili ključne riječi nisu više potrebni, već da bi se u slučaju elektroničke građe oni morali dodavati u fazi stvaranja građe.

Drugi stručnjaci²⁶⁰ smatraju da bi bibliografsku kontrolu nacionalne produkcije mogli olakšati kad bi nakladnici i bibliografska središta sudjelovali tako da sami autori i izdavači elektroničkih dokumenata kreiraju²⁶¹ metapodatke koji su namijenjeni identifikaciji i lociranju te građe. Jedno od rješenja je primjena Osnovnoga dublinskog skupa elemenata metapodataka (*Dublin Core*).

S elektroničkom građom otvaraju se nova područja, kao što su upravljanje autorskim pravima, osiguravanje autentičnosti građe, dugoročna pohrana njihova pristupa, upravljanje građom s više formata, trajnost URL-a itd. Po Lupoviciju, za mrežnu je građu najvažnije zabilježiti sve tehničke i druge specifičnosti građe u trenutku njihova ulaska u arhivski sustav, uključujući i upute za robota za prikupljanje i moguću suradnju s autorima građe. U slučaju stvaranja nacionalne bibliografije mrežne građe, ona bi izgledala kao popis URL-a s indeksom imena ili predmetnim indeksom.

Među europskim nacionalnim knjižnicama očito postoji dvojba u vezi s bibliografskim opisom mrežne građe i njezinim uključivanjem u nacionalne bibliografije. Wiggins²⁶² je analizirao nacionalne bibliografije 32 europske nacionalne knjižnice i bibliografske agencije. Njegovo je istraživanje pokazalo sljedeće:

- u šesnaest zemalja mrežna se građa prikuplja robotom za prikupljanje građe,

²⁶⁰ Klarin, S. ; Murati, T. Identifikacija, odabir, obrada i osiguravanje dostupnosti mrežnih publikacija u kontekstu nacionalne bibliografske kontrole // 4. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u kontekstu globalne informacijske infrastrukture. Zagreb : Hrvatsko knjižničarsko društvo, 2001. Str. 41-56.

²⁶¹ Klarin, Pigac i Pavečić su u okviru istraživanja hrvatskih serijskih publikacija na mreži utvrdili da su samo 33.3% od 152 istraženih naslova uključili metapodatke u zaglavlje HTML. Vidi djelo: Klarin, S. ; Pigac, S. ; Pavelić D. Croatian remote access electronic serials: results of a survey. // *International Cataloguing & Bibliographic Control*, 30, 4(2001), str. 70-72.

²⁶² Wiggins, B. IFLA survey on inclusion of electronic resources in national bibliographies. // *World Library and Information Congress: 71th IFLA General Conference and Council "Libraries - A voyage of discovery"*, August 14th-18th 2005, Oslo, Norway.

- jedanaest zemalja uključuje tu građu u svoju nacionalnu bibliografiju,
- jedanaest zemalja prikuplja mrežnu građu po nacionalnoj domeni,
- dvanaest zemalja selektivno prikuplja građu u okviru nacionalne domene,
- devet zemalja prikuplja građu izvan nacionalne domene (*.net, *.org, *.com i druge),
- sedam zemalja prikuplja samo publikacije koje su nastale u digitalnom obliku,
- sedam zemalja prikuplja samo građu, koja je besplatno dostupna,
- dvije zemlje prikupljaju građu koja nije besplatna,
- dvije zemlje prikupljaju građu koja je dostupna s ograničenjima

Zanimljiv je podatak da dvadeset institucija uključuje u nacionalnu bibliografiju svu mrežnu građu s cijelim bibliografskim opisom po uzoru američku Kongresnu²⁶³ i australsku Nacionalnu knjižnicu.²⁶⁴ Pritom se osam njih koristi angloameričkim katalogizacijskim pravilima (AACR), a trinaest se upravlja po standardima ISBD. Čak šest institucija ima normativnu kontrolu autorskih odrednica.

U okviru mrežne građe u nacionalnim bibliografijama najčešće je obuhvaćena građa koja nalikuje tradicionalnim oblicima publiciranja (elektroničke knjige, elektroničke revije ...), ali ima i slučajeve u kojima su uključeni blogovi, naslovnice i mrežne stranice iz cijele nacionalne domene.²⁶⁵

²⁶³ U Kongresnoj knjižnici koriste se programskom opremom konzorcija OCLC (CORC odnosno OCLC Connect). Bibliografski opis uključuju u svoj integralni knjižnični sustav (*Library of Congress's Integrated Library System*). Ne katalogiziraju sve pojedinačne mrežne stranice, već tematske skupine stranica ili mrežna mjesta. Pritom se koriste formatom MODS. U strukturu formata za katalogizaciju dodali su polje 583, u kojem je označeno da je *Library of Congress* već arhivirala taj dokument, i polje 856, u koje je unesen PURL do arhiva (pomoću programa Handle²⁶³), u slučaju da poslužitelj prestane s radom. CORC omogućava preslikavanje tako što dođemo do određenoga dokumenta na mreži i pritiskom na tipku automatski preuzmemo metapodatke iz mrežnog dokumenta u HTML. Moguće je definirati format u kojem želimo smjestiti metapodatke ili označiti stranice koje želimo da se automatski obrade. CORC automatski klasificira dokument u Dewey. Omogućava i normativnu kontrolu autorskih imena i povezan je s *OCLC Authority Files*. Podupire više formata.

²⁶⁴ Katalogizacija u PANDORA-i obavlja se u formatu MARC 21 (u polje 856 unose URL poveznice) dok deskriptivne metapodatke od nakladnika dobivaju u formatu Dublin Core. Koriste se i strukturnim, tehničkim i administrativnim metapodacima. Dokumenti su smješteni u bazi podataka, koju su upravo za tu svrhu u nacionalnoj knjižnici osnovali već 1997. godine.

²⁶⁵ Jacobsen, G. Universal Bibliographic Control and International MARC Core Programme (UBCIM) : International Conference on National Bibliographic Services Copenhagen, 25-27 November 1977: COVERAGE. // ICNBS, Copenhagen, 25-27 November 1998. URL: <http://www.ifla.org/VI/3/icnbs/fina.htm>. (15.12.2006.)

Pojava elektroničkih publikacija primorala je IFLA-u da preispita stara međunarodna katalogizacijska pravila, takozvana Pariška načela iz 1961. godine. Katalogizacijska praksa se s vremenom mijenjala i 1977. godine dobili smo nov standard za opis neknjižne građe, dok je 1988. godine je nastao prvi standard za računalne datoteke (ISBD(ER)).²⁶⁶ Zbog novih katalogizacijskih uvjeta IFLA je 2003. godine započela s pripremama novoga međunarodnoga katalogizacijskoga pravilnika, čiji je cilj uvođenje standarda za opis bibliografskih i normativnih zapisa u knjižničnim katalozima, da bi se olakšala izmjena podataka u međunarodnom prostoru. Nov međunarodni katalogizacijski pravilnik značio bi modernizaciju pariških načela, uveo bi terminologiju i koncept FRBR te uskladio važeća pravila u vezi s katalogizacijom osobnih imena, korporacija, jedinstvenih naslova građe, zbirke i višerazinskom katalogizacijom.²⁶⁷ Tako bi nova načela zamijenila i dopunila pariška načela tako da budu upotrebljiva za sve vrste građe, i to ne samo za tekstove, već i za sve aspekte bibliografskih i normativnih zapisa koji se koriste u knjižničnim katalozima. Nova načela temeljila bi se na konceptu FRBR (*Functional Requirements for Bibliographic Records*)²⁶⁸ i FRANAR (*Functional Requirements for Authority Records*).²⁶⁹

Moguće je da će nova međunarodna pravila i promjene koje se uvode u praksu katalogizacije dijelom riješiti i problem obradbe mrežne građe. Dijelom, jer se još uvijek operira konceptom materijalne građe. Pokušava se prikupiti djelo kao apstraktni pojam, ali u okviru postojeće konvencionalne građe. Imamo elektroničke knjige, serijske publikacije, zbornike itd. Opis građe morao bi se usredotočiti na entitet *dijela* i *izraza*, jer entitet *manifestacija* nije toliko relevantan, a entitet *izdanje* je neupotrebljiv. Iako Nuys i Albertsen pokazuju suprotno, entiteti manifestacija i

²⁶⁶ Popović-Bošković, G. ; Fortuna, B. Novi mednarodni katalogizacijski pravilnik : Korporativna značnica med idejo o enotni in univerzalno kontrolirani obliki // Organizacija znanja, 9, 2(2004). URL: http://home.izum.si/COBISS/OZ/2004_2/html/clanek_02.html. (15.12.2006.)

²⁶⁷ Izjava o mednarodnih katalogizacijskih načelih: osnutek, sprejet na 1. srečanju strokovnjakov za izdelavo mednarodnega katalogizacijskega pravilnika u okviru IFLE, Frankfurt, Nemčija, 2003 / prevele Kavčič, I. i Kalčić, D. URL: http://www.ddb.de/standardisierung/pdf/statement_slovene.pdf. (15.12.2006.)

²⁶⁸ Functional Requirements for Bibliographic Records : Final Report // IFLA Study Group on the Functional Requirements for Bibliographic Records. München : K.G. Saur, 1998. (UBCIM Publications - New Series Vol. 19). - URL: <http://www.ifla.org/VII/s13/frbr/frbr.pdf>. (15.12.2006.)

²⁶⁹ Functional Requirements for Authority Records: A conceptual model // IFLA UBCIM Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). Osnutak, 15.6.2005. URL: <http://www.ifla.org/VII/d4/Franar-Conceptual-M-Draft-e.pdf>. (15.12.2006.)

izdanje relevantni su kod dinamičnih mrežnih publikacija.²⁷⁰ Činjenica je da model FRBR još nije upotrijebljen u knjižničnoj praksi²⁷¹ i proći će dosta vremena do njegove implementacije, kao i do uvođenja u praksu novih međunarodnih katalogizacijskih načela.

Godine 2002. revizijom angloameričkih katalogizacijskih pravila (AACR2)²⁷² dobili smo novu podjelu građe na neomeđenu²⁷³ i integrirajuću. Njihova podjela uzeta je u obzir u reviziji standarda ISBD(S) - ISBD(CR). Nova podjela pomogla je pri uključivanju publikacija na mreži. Međutim, podjela je vrlo opća i stvara nejasnoće kad je u pitanju mrežna građa. Novom podjelom dodaje se opis "sive literature" (osuvremenjivana publikacija koja ima uvez sa slobodnim listovima), koja je prije bila isključena. Zanimljivo je i to da se opis neomeđene građe temelji na prvom zapisu, a opis integrirajuće građe na zadnjoj iteraciji. U SAD-u su nakon revizije AACR2 nastavili s revizijom MARC21, LCRI (*Library of Congress Rule Interpretation*) i konačno ISSN. Takvo je usklađivanje olakšalo izmjenu podataka među različitim institucijama.²⁷⁴

Nakon prijevoda standarda u slovenski jezik NUK je u suradnji s IZUM-om također pristupio usklađivanju slovenskoga UNIMARC-a, COMARC-a. Sad se već može uključiti neomeđena građa, u okviru koje se po ISBD(CR) obrađuje sva integrirajuća građa (neomeđena i zaključena).

²⁷⁰ Nuys, C. van ; Albertsen, K. Identification of network accessible documents: problem areas and suggested solutions // 3rd ECDL Workshop on Web Archives, August 21st, 2003 Trondheim, Norway. URL: <http://bibnum.bnf.fr/ecdl/2003/index.html>. (15.12.2006.)

²⁷¹ Nuysova i Albertsen eksperimentiraju s modelom FRBR za mrežnu građu. Model bi bio primjenljiv kod statičnih publikacija, naprimjer (str. 15) ako Norveški odbor za zdravlje objavi upute za upotrebu antibiotika u bolnicama, te upute možemo nazivati razinom *djela* jer nose zajednički stvarni naslov. Pri odabiru nakladničkoga izdanja spuštamo se na razinu *izraza*, pri odabiru formata dolazimo do razine *manifestacije* i na kraju, ako primjerak prilagodimo našim potrebama, onda je to razina *izdanja*. Međutim, kod dinamičnih bi publikacija URL adresa koju bismo dobili nakon pretraživanja, naprimjer po bibliografskoj bazi, predstavljala bi jednu manifestaciju te baze, a ako bismo preslikali dokument koji se neprekidno mijenja, onda bismo snimkom dobili njegovo izdanje. Pored opisanog iskustva u Norveškoj Gambari i Guerrini opisuju primjer australskog OPAC-a AustLit (*Australian literature gateway*), koji primjenjuje relacijski model FRBR Vidi: Gambari, S., Guerrini, M. Definire e catalogare le risorse elettroniche. Milano : Editrice Bibliografica, 2002.

²⁷² *Anglo-American Cataloguing Rules*. Second Edition, 2002 Revision / prepared under the direction of the Joint Steering committee for Revision of AACR, a committee of the American Library Association [et al.]. Chicago : American Library Association ; Ottawa : Canadian Library Association ; London : Chartered Institute of Library and Information Professionals, 2004.

²⁷³ U tu se skupinu ubrajaju elektroničke verzije tiskanih serijskih publikacija, elektroničke serijske publikacije, elektronske novine, bibliografske baze podataka, elektroničke diskusijske skupine, mrežna mjesta i digitalne knjižnice (Pênkavová, P. *E-serials processing methodology*. // Newsletter of the IFLA Section on Serial Publications. 39(2001), str. 12.

²⁷⁴ Popović Bošković, G. ; Fortuna, B. ISBD(CR) : Izzivi katalogizacije virov u nadaljevanju. // Organizacija znanja 8, 2(2003).

Bibliografska obrada mrežne građe još uvijek ostavlja otvorena sljedeća pitanja:

- je li bibliografski opis mrežne građe, za koji se zauzimaju različite međunarodne stručne skupine, dovoljan za opis mrežne građe;
- je li potreban pun bibliografski opis za svu građu ili je dovoljan opis petnaest elemenata iz sheme Osnovnog dublinskog skupa elemenata metapodatka (*Dublin Core*);²⁷⁵ ne smijemo zaboraviti da se sami elektronički dokumenti mogu indeksirati i biti locirani pomoću mrežnih pretraživača;
- kako odrediti glavni izvor podataka za bibliografski opis mrežne građe - kod većine integrirajuće građe nema podataka o autoru ili godini publikacije, ili se ne zna koji je glavni stvarni naslov jedinice građe;
- kako riješiti problem bibliografske kontrole i osigurati autentičnost mrežne građe;
- kako pratiti i različite inačice iste građe te kako postupiti s njima (različiti formati, različite verzije itd.); čak i autori naslovnica često mijenjaju njihov oblik i sadržaj jer se još uvijek teško snalaze u dizajniranju i definiranju naslovnice; ta nestabilnost i promjenljivost mrežne građe može otežati proces katalogizacije;²⁷⁶
- trebalo bi razmotriti uvođenje nove tipologije dokumenata na mreži, koji bi dopunili bibliografski opis.

U praksi je običaj da se za građu čija struktura i forma odgovara formalnoj nakladničkoj djelatnosti koriste jednake metode obradbe kao i kod materijalne građe, sa dodatnim informacijama o načinu pristupa i formatima (po UNIMARC-u ili MARC-u). Sve promjene, u slučaju da ne utječu na sadržaj, vodile bi se u napomenama istog zapisa. U procesu arhiviranja dodali bi se ostali metapodatci, tehnički, administrativni ili za dugoročnu pohranu.

Problem nastaje kod automatski prikupljene mrežne građe. Rješenje bi bilo automatsko kreiranje osnovnoga bibliografskoga opisa. Tehnički i administrativni metapodatci te metapodatci za dugoročnu pohranu jednostavno se automatski generiraju. Postavlja se pitanje je li potreban dodatni bibliografski opis i ako jeste, kako će robot automatski prepoznati relevantne

²⁷⁵ Hakala, J. *Internet metadata and library cataloguing*. // *International Cataloguing & Bibliographic Control*. 28, 1(1999), str.21-25, predlaže upotrebu Dublin Core, navodeći argument da postoje konverzijske tabele u MARC-u.

²⁷⁶ Klarin, S. ; Murati, T. Identifikacija, odabir, obradba i osiguravanje dostupnosti mrežnih publikacija u kontekstu nacionalne bibliografske kontrole // 4. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u kontekstu globalne informacijske infrastrukture. Zagreb : Hrvatsko knjižničarsko društvo, 2001. str. 48-49.

metapodatke. Kod većine mrežnih stranica napisanih u HTML-u ili XML-u to je moguće, dok je kod drugih, naprimjer dinamičnih stranica, to problem. U IIPC-u pokušavaju riješiti taj problem uvođenjem formata WARC.

Uz svim spomenutim raspravama nude se rješenja, koja se temelje na mogućnostima informacijske tehnologije i alternativa su ili dopuna postojećim bibliografskim katalozima, posebice za novonastale tipove elektroničke građe. *World Wide Web Consortium* je u tu svrhu 1999. godine razvio RDF (*Resource Description Framework*)²⁷⁷ model koji se koristi za stvaranje metapodataka o semantičkim odnosima u elektroničkom dokumentu i omogućava preciznija semantička pretraživanja. Obično se ugrađuje u sintaksu XML. Nažalost, njegova je upotreba ostala ograničena na zatvorene istraživačke skupine i projekte na institutima u okviru semantičkih analiza i *text mining* metoda. Ove metode pružaju mogućnosti različite vizualizacije građe i otkrivanja međusobnih relacija jedinica različite građe u svjetskoj mreži. Još nije došlo do praktične primjene tih metoda u knjižnicama ili su još u eksperimentalnoj fazi.²⁷⁸

Sve su češći portali koji organiziraju pristup mrežnoj građi. U akademskim okruženjima oni se kombiniraju s predmetnim indeksima i različitim anotacijama koje omogućavaju lakši odabir relevantne mrežne građe. To nikako ne može u potpunosti zamijeniti bibliografski opis, međutim može biti dodatan alat za pronalaženje građe u web arhivu.²⁷⁹

²⁷⁷ URL: <http://www.w3.org/RDF/>. (15.12.2006)

²⁷⁸ Jørgensen, P. H. navodi primjer Danskoga centra za knjižnice, gdje su kombinirali IFLA FRBR model s RDF u okviru njihova sustava za katalogizaciju Visual Cat. Ukratko o tome u Jørgensen, P. H. Practical application of FRBR and RDF. // *Le risorse elettroniche : definizione, selezione e catalogazione = Electronic resources. Definition, selection and cataloguing. Proceedings of the International Conference, Rome, November 26-28, 2001 / edited by Guerrini, M. Milano : Bibliografica, 2002. str. 623.*

²⁷⁹ Weston citira Michaela Gormana kako bibliografska obradba ima vrlo ograničene mogućnosti u računalnom okruženju. U tom istom djelu spominje tvrdnju Alfreda Serraia, da je dosad bibliografska struka pokušala samo prenijeti postupke iz tradicionalne građe u elektroničko okruženje. Ako je funkcija kataloga otkrivanje fondova, onda te postupke treba svakako prilagoditi novim medijima i alatima. Više o tome u Weston, P. G. Between bibliographic and intellectual control. Some reflections on electronic bibliographic systems. // *Le risorse elettroniche : definizione, selezione e catalogazione = Electronic resources. Definition, selection and cataloguing. Proceedings of the International Conference, Rome, November 26-28, 2001 / izdao Guerrini, M. Milano : Bibliografica, 2002. - str. 609-621.*

9. DUGOROČNA POHRANA PUBLIKACIJA NA MREŽI

Problematika dugoročne pohrane više je od deset godina na dnevnom redu mnogih institucija koje grade elektroničke arhive. Tehnologija se mijenja vrlo brzo, a znanje koje je vezano uz tu tehnologiju, vezano je na generaciju stručnjaka. Ako se navrijeme ne planiraju i ne uspostave uvjeti za dugoročnu pohranu te građe, moglo bi se dogoditi da za nekoliko godina, s uvođenjem nove tehnologije ili novih operacijskih sustava, tu građu nećemo moći pročitati.

Na taj su problem među prvima upozorile američke svemirske agencije, koje su zbog velikih količina elektroničkih dokumenata, dobivenih iz satelitskih i ostalih izvora, morale hitno potražiti rješenje za njihovu dugoročnu pohranu. Kao posljedica suradnje NASA-e s Međunarodnom organizacijom za standardizaciju (ISO) 2003. godine prihvaćen je prvi standard za građu digitalnih arhiva, ISO 14721, poznatiji kao referentni model OAIS.²⁸⁰

U Europi se dugoročnoj pohrani elektroničke građe posvećivalo više pažnje u okviru projekta NEDLIB (*Networked European Digital Library*). Cilj projekta bio je uspostavljanje depozitarnoga sustava za dugoročnu pohranu elektroničkih publikacija. Njihov se sustav temeljio upravo na prije spomenutom referentnom modelu, koji tada još nije bio prihvaćen kao standard. Projekt NEDLIB završen je 2000. godine, iako je njegov model arhivskoga sustava, koji je razvijen u okviru projekta, još uvijek koristan, kao i svi izvještaji i istraživanja koji su objavljeni na njihovoj polaznoj stranici. Uvjerljiv je dokaz o uspješnosti implementacije NEDLIB-ova modela još uvijek aktivan depozitarni arhiv elektroničkih publikacija DNEP, u Kraljevskoj knjižnici u Nizozemskoj.

Na dugoročnu pohranu elektroničkih zapisa uvelike su utjecali američka Skupina istraživačkih knjižnica (*Research Libraries Group* - RLG) i američki konzorcij OCLC (*Online Computer Library Center*). Već su 1994. godine Radna skupina RLG za arhiviranje digitalnih informacija (*RLG Taks Force on Archiving of Digital Information*) i Komisija za dugoročnu pohranu i pristup (*Commission on Preservation & Access*) iz Washingtona istaknuli nužnost određivanja uvjeta

²⁸⁰ Reference Model of an Open Archival Information System (OAIS) : Recommendation for space data system standards : CCSDS 650.0-B-1.Blue Book, January 2002 // CCSDS (Consultative Committee for Space Data Systems). Washington : CCSDS Secretariat. URL: http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html. (21.9.2004)

koji bi morali ispunjavati pouzdani digitalni arhivi. U njihovu izvještaju iz 1996. godine između ostalog se navodi da su za dugoročnu pohranu digitalne građe potrebni odgovarajuća infrastruktura za potporu distribuiranoga sustava digitalnih arhiva dovoljan broj pouzdanih organizacija, koje bi bile sposobne čuvati, migrirati i omogućiti dostup do digitalnih zbirka te uvođenje prakse dodjeljivanja certifikata, koji bi povećali pouzdanost arhiva.²⁸¹

Godine 2000. RLG i OCLC započeli su sa suradnjom i kao rezultat toga dvije godine poslije RLG je objavio izvještaj: *Trusted digital repositories: attributes and responsibilities*.²⁸² Publikacija je vrlo značajna jer obrađuje svojstva i odgovornosti pouzdanih digitalnih arhiva. (*trusted digital repositories*). Među autorima su poznati stručnjaci iz područja dugoročne pohrane elektroničke građe. Catherine Lupovici, voditelj Odjela za digitalnu knjižnicu u francuskoj Nacionalnoj knjižnici, sudjelovala je u europskom projektu NEDLIB (*Networked European Deposit Library*),²⁸³ čiji je cilj bio razvoj depozitarnoga sustava za dugoročnu pohranu digitalnih publikacija. Njezin doprinos u projektu je razvoj modela metapodataka za dugoročnu pohranu u digitalnom repozitoriju.²⁸⁴ Deborah Woodyard, tadašnja koordinatorica za dugoročnu pohranu digitalne građe u *British Library*, vrlo je aktivna u RLG-u, i to na području definicije metapodataka za dugoročnu pohranu. Colin Webb, direktor Odjela za dugoročnu pohranu u australskoj Nacionalnoj knjižnici, autor je djela *Guidelines for the preservation of digital heritage*²⁸⁵, iz 2003. godine, (u nastavku *Smjernice*), koje je nastalo po narudžbi UNESCO-a. Spomenute *Smjernice* također su velik doprinos dugoročnoj pohrani. Nastale su kao posljedica UNESCO-ove povelje o dugoročnoj pohrani digitalne baštine, koja je prihvaćena na 32. zasjedanju Generalne Konferencije UNESCO-a, u listopadu 2003., i obrađuje sve aspekte gradnje digitalnih arhiva.

²⁸¹ Garret, J. Waters, D. *Preserving Digital Information. Report of the Task Force on Archiving of Digital Information* (1996). Commission on Preservation and Access and RLG. URL: <http://www.rlg.org/legacy/ftpd/pub/archtf/final-report.pdf>. (7.2.2004)

²⁸² *Trusted Digital Repositories: Attributes and responsibilities : an RLG-OCLC Report*. Mountain View, CA : RLG. URL: <http://www.rlg.org/longterm/repositories.pdf>. (21.9.2004)

²⁸³ Domaća stran projekta <http://www.kb.nl/coop/nedlib/> (21.9.2004.)

²⁸⁴ Lupovici, C. ; Masanès, J. *Metadata for Long Term Preservation*. Den Haag : Koninklijke Bibliotheek, 2000. (NEDLIB Report series ; 2).

²⁸⁵ *Guidelines for the preservation of digital heritage* (2003) / prepared by the National Library of Australia. Paris : UNESCO. <http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>. (21.9.2004)

Među novijim europskim projektima valja spomenuti projekt reUSE²⁸⁶ (2004.-2006. - program eContent) čiji je cilj bio u nacionalnim knjižnicama Estonije i Austrije te u sveučilišnim knjižnicama u Innsbrucku, Grazu i knjižnici Sveučilišta Humboldt u Berlinu izgraditi pouzdane digitalne arhive za digitalne oblike publikacija koje su bile objavljene samo u tiskanom obliku. U okviru toga projekta NUK je vodio evaluacijsku radnu skupinu, koja je trebala ocijeniti digitalne arhive koje su drugi partneri gradili.²⁸⁷ Prvu fazu projekta posvetili smo razvoju uvjeta u institucijama za dugoročnu pohranu elektroničke građe u smislu izvještaja RLG i OCLC. Pritom smo pobliže pratili organizacijski, tehnološki i korisnički aspekt izgrađenih arhiva, koji moraju biti tijesno povezani jer se međusobno isprepliću. Izgradnja pouzdanoga digitalnoga arhiva mora biti dio integralnoga pristupa dugoročnoj pohrani elektroničke građe i zato mora biti usklađena sa svim djelatnostima na tom području.

Već spomenute UNESCO-ove *Smjernice* među prvima su odredile praksu dugoročne pohrane elektroničke građe, i to bez obzira na to je li ona prikupljena na mreži ili iz drugih izvora. Glavni autor djela, Colin Webb, radi u australskoj Nacionalnoj knjižnici. Poznat je njihov krajnje strog selektivni pristup prikupljanju mrežne građe. *Smjernice* uzimaju u obzir sve spomenute karakteristike i odgovornosti pouzdanih digitalnih arhiva. Sadrže preporuke za izgradnju digitalnih arhiva te opisuju postupke i aspekte na koje moramo obratiti pažnju. Ti postupci i aspekti su sljedeći:

a) Preuzimanje odgovornosti za dugoročnu pohranu digitalne građe

Organizacija koja preuzima zadatak dugoročne pohrane elektroničke građe mora biti svjesna svojih odgovornosti koje time preuzima. U *Smjernicama* se navodi da te odgovornosti mogu biti podijeljene među različitim institucijama (naprimjer više knjižnica, arhiva ili muzeja), ili mogu biti ograničene na određeno vremensko razdoblje. Važno je da te institucije imaju iskustva s dugoročnom pohranom kulturne baštine. Ne isključuje mogućnosti i komercijalnih repozitorija. Moraju voditi računa i o kontinuitetu čuvanja.

²⁸⁶ Project reUSE: <http://reuse.uibk.ac.at/>. (21.9.2004)

²⁸⁷ Izvještaj o evalvaciji je dostupan na URL: <http://www.uibk.ac.at/reuse/documents/>. (21.9.2004)

b) Kriteriji odabira digitalne građe

Prilikom preuzimanja odgovornosti važno je definirati građu za koju smo odgovorni. O odabiru je već bilo riječi u drugom poglavlju. *Smjernice* daju prednost građi koja je kreirana u elektroničkom obliku i sadržajima koji predstavljaju kulturnu i znanstvenu baštinu naroda. Ustanovama se prepušta da same odrede svoje kriterije, ovisno o njihovu sustavu za arhiviranje. Također se preporučuje određivanje pravila odabira na samom početku prikupljanja građe, čime bi se riješile sljedeće dvojbe:

- kako postupiti kada dobijemo istu građu na različitim medijima ili formatima;
- kako postupiti kada dobijemo noviju verziju (izdanje) elektroničke građe, koja je identična prvoj;
- kako možemo ustanoviti vjerodostojnost originalne digitalne građe.

Pri odabiru građe pomoći će ako pokušamo predvidjeti potencijalne korisnike u budućnosti, kao i mogućnost osiguravanja svih pravnih dopuštenja za njihovu dugoročnu pohranu. Troškovi njezina prikupljanja, arhiviranja i održavanja ne bi smjeli biti veći od njezine vrijednosti.

c) Suradnja s autorima i nakladnicima digitalne građe

Smjernice ističu veliku važnost suradnje digitalnih arhiva s autorima, odnosno nakladnicima elektroničke građe, jer je na taj način lakša izmjena informacija o svim sudionicima u procesu publikacije, kao i informacija koje su potrebne za dugoročnu pohranu građe. I u početnim fazama moguće je dogovarati se o formatima i oblicima, standardizirati postupke pri njihovu objavljivanju, utvrditi ciljanu publiku itd. Što je najvažnije, tom suradnjom osigurana je i cjelovitost i autentičnost građe.

d) Uređivanje aspekata autorskoga prava

Smjernice navode tri različite vrste autorskih prava, koje moramo osigurati za dugoročnu pohranu elektroničke građe:

- pravo na prikupljanje građe (preslikavanje građe u arhiv)
- pravo pristupa (preslikavanje građe na radni prostor računala u mreži)
- pravo na upotrebu strategije za dugoročnu pohranu (kopija građe, primjena različitih intervencija nad građom - konverzije, migracije itd.)

Dok nacionalnim knjižnicama prvo pravo određuje Zakon o obveznom primjerku, za druga dva, posebice ako je u pitanju komercijalna građa, potrebno je sklapanje ugovora ili dobivanje licence.²⁸⁸

e) Odabir metodologije dugoročne pohrane elektroničke građe

U dobrom programu dugoročne pohrane vrlo je da prije unosa građe u arhiv odredimo moguće strategije za njezinu dugoročnu pohranu. Odabir strategije ovisi o dopuštenju vlasnika autorskih prava i o vrijednosti same građe. Elektronička građa, kao što ćemo objasniti u posebnom poglavlju, sastoji se iz više slojeva. U trenutku arhiviranja moramo ustanoviti koji aspekt želimo sačuvati za budućnost.

f) Usklađenost arhivskoga sustava s referentnim modelom OAIS

Usklađenost s referentnim modelom Otvorenoga arhivskoga informacijskoga sustava (OAIS) jedan je od uvjeta koji zahtijevaju gotovo svi autori na području dugoročne pohrane elektroničke građe. Zbog svoje važnosti bit će detaljnije prikazan u posebnom poglavlju.

g) Odabir primjernih metapodataka

Metapodatci su veoma važni za lokalizaciju i administriranje građe te u procesima dugoročne pohrane, jer nose informaciju o građi i time joj omogućavaju pristup. Mnoga su istraživanja posvetila puno pažnje metapodacima. O tome će također biti govora u posebnom poglavlju.

Među važnijim inicijativama na području dugoročne pohrane elektroničke građe danas možemo nabrojiti mrežni portal australske Nacionalne knjižnice PADI (<http://www.nla.gov.au/padi/index.html>), koji redovito objavljuje važne informacije sa svih područja dugoročne pohrane elektroničkih publikacija iz cijelog svijeta, te prije spomenuti konzorcij *International Internet Preservation Consortium* (<http://netpreserve.org/>), gdje nalazimo informacije o standardima i mogućnost preslikavanja softvera za prikupljanje mrežne građe. Među nacionalnim inicijativama vrijedi spomenuti njemački projekt KOPAL (<http://kopal.langzeitarchivierung.de/index.php.en>), koji je uzor nacionalnoga rješavanja

²⁸⁸ 1999. godine su Konferencija europskih nacionalnih knjižnica (CENL) i Federacija europskih nakladnika (FEP) pripremili dokument *"Code of practice for the voluntary deposit of electronic publications"* u kome su odredili i osnove za suradnju knjižnica sa nakladnicima i autorima kod depozita elektroničke građe.

problematike dugoročne pohrane elektroničkih publikacija, zatim NESTOR (<http://www.langzeitarchivierung.de/index.php?newlang=eng>), ili britanski *Digital Curator Centre* (<http://www.dcc.ac.uk/about/>), gdje također nalazimo portal sa zanimljivim publikacijama iz tog područja, te *Digital Preservation Coalition* (<http://www.dpconline.org/graphics/>) koji je primjer suradnje između različitih vrsta institucija na tom istom području.

Uz spomenuti izvještaja, RLG i OCLC osnovali su ekspertnu mrežu PREMIS (*PREservation Metadata Implementation Strategies*) u svrhu određivanja minimalnoga izbora metapodataka za dugoročnu pohranu digitalne građe.²⁸⁹

Austrijska konzultantska tvrtka *ARGE Krimbacher, Neuhauser and Vogl* u okviru već spomenutoga europskoga projekta reUSE²⁹⁰ 2005. godine izvela je istraživanje²⁹¹ o različitim aspektima dugoročne pohrane u 330 nacionalnih, sveučilišnih i istraživačkih knjižnica u 25 europskih zemalja. Istraživanje je pokazalo da 76% knjižnica daje najveću moguću važnost dugoročnoj pohrani, 91% je mišljenja da je to jedan od važnijih zadataka knjižnica, 38% ispitanih knjižnica izjavilo je da ima digitalni arhiv za dugoročnu pohranu elektroničke građe, a isti postotak knjižnica namjerava ga uskoro izgraditi. Zanimljivo je da su na pitanje za koliko godina misle da će biti dostupna elektronička građa (ponuđeni su odgovori za 5, 20, 50, 100 i 500 godina), najviše optimizma pokazale knjižnice iz Velike Britanije, Irske i skandinavskih država. One vjeruju da će u svakom slučaju ta građa biti dostupna za 20 ili čak za 50 godina. Postoji srednja vjerojatnost da će biti dostupna i za 100 godina, a postoji i mogućnost za 500 godina. Ovakav optimizam rezultat je bogate prakse na području dugoročne pohrane. Takvo je trenutačno stanje u Europi.

²⁸⁹ Opis projekta PREMIS na URL: <http://www.oclc.org/research/projects/pmwg/>. (7.2.2005.)

²⁹⁰ Project reUSE: <http://reuse.uibk.ac.at/>. (15.12.2006.)

²⁹¹ Krimbacher, M. ; Neuhauser, M. ; Vogl, M. Survey on the long-term preservation of digital documents in European Libraries. Innsbruck: Krimbacher Neuhauser & Vogl, 2005. http://1.1.1.1/270413892/282738784T061107124511.txt.binXMyS0dapplication/pdfXsysM0dhttp://www.uibk.ac.at/reuse/docs/d_6.7study_european_digital_repositories.pdf (15.12.2006.)

9.1 Strategije za dugoročnu pohranu elektroničke građe

Strategija za dugoročnu pohranu elektroničke građe tijesno je vezana uz različite razine od kojih se sastoje elektronički dokumenti ili digitalni objekti²⁹². U nastavku ću prikazati strukturu elektroničke građe, a zatim različite strategije njihove dugoročne pohrane.

9.1.1 Struktura elektroničke građe

Postoje četiri osnovne razine^{293,294} od kojih se sastoje elektroničke publikacije.²⁹⁵ (vidi **sliku br. 10**):

- **Materijalna razina** - to je digitalni zapis ili sama oznaku na nekom materijalnom nositelju, koji može biti CD-ROM, disketa, magnetna vrpca ili drugo;
- **Logička razina** - to je sintaksa ili programski jezik u kojem je napisan, odnosno sačuvan zapis; pri prijelazu u novije verzije računalnih programa može se sasvim promijeniti, a to se događa i pri konverzijama u druge aplikacije;
- **Konceptualna razina** – ona nam omogućava pristup sadržajima elektroničke publikacije; na toj razini doznajemo više o sadržaju, naprimjer je li u pitanju tabela, slika, tekst ili zvučna snimka; nije nam važno je li tekst u formatu PDF ili u MS Wordu, važno je da ga možemo pročitati;
- **Razina posebnih karakteristika elektroničke građe** - to je skup svih specifičnih svojstava koje čine elektroničku jedinicu građe time što je i sam korisnik može primijetiti; naprimjer, za stručnjaka koji želi analizirati sliku muzejskoga izložka veoma je važna njezina rezolucija i raspon boja; isto tako, ako želimo promijeniti neki tekst, nije nam svejedno je li on dostupan

²⁹² Sustavna analiza operira s konceptom digitalnoga objekta, kojim se označava bilo koja završena intelektualna cjelina elektroničkoga zapisa. Digitalni objekt može biti elektronička knjiga, elektronička slika, digitalni video, neka druga vrsta zapisa, odnosno kombinacija zapisa. Može se sastojati od više datoteka ili samo od jedne. Mnoge metode i koncepti iz sustavne analize primjenjuju se u bibliotekarstvu, jer se oba područja bave nastankom elektroničkih publikacija. Zato sve češće u literaturi nalazimo opći naziv digitalnoga objekta za navođenje bilo koje jedinice elektroničke građe.

²⁹³ Thibodeau, K. L. *Overview of technological approaches to digital preservation and challenges in coming years*. URL: <http://www.clir.org/pubs/reports/pub107/thibodeau.html>

²⁹⁴ Lupovici, C. ; Masanès, J. *Metadata for the long-term preservation of electronic publications*. The Hague : NEDLIB Consortium, 2000. (NEDLIB Report Series; 2), URL: <http://1.1.1.1/315824780/298360064T061215155831.txt.binXMysM0dapplication/pdfXsysM0dhttp://nedlib.kb.nl/reults/NEDLIBmetadata.pdf>. (15.12.2006.), str. 10.

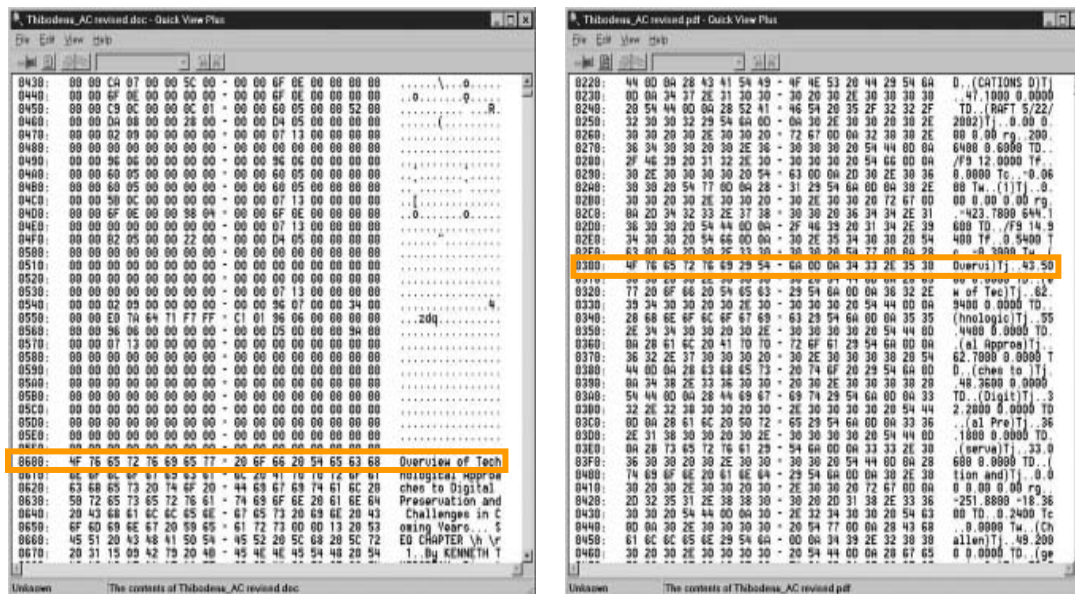
²⁹⁵ Thibodeau te razine naziva digitalnim objektima, za svaku razinu ili perspektivu gledanja elektroničke publikacije po jedan digitalni objekt.

u PDF formatu, JPEG formatu ili u MS Wordu jer svaki od tih formata ima različite funkcije itd.



Slika br. 10: Struktura elektroničke građe

Thibodeau ilustrira dvije različite logičke razine jednoga dokumenta. (slika br. 11)



Slika br. 11: Dvije različite logičke razine jednoga dokumenta. (Izvor: Thibodeau, K. L. Overview of technological approaches to digital preservation and challenges in coming years. URL: <http://www.clir.org/pubs/reports/pub107/thibodeau.html> - crvenim je dodatno označen početak teksta)

S lijeve je strane prikazan dokument u MS Wordu, a s desne u PDF-u. Bojom je označen početak teksta.

Pri odabiru strategije za dugoročnu pohranu elektroničkih publikacija vrlo je važno znati što želimo sačuvati u njima. Želimo li održati njihovu prvobitnu logičku strukturu ili pak mislimo da je skup svih njihovih funkcionalnosti i karakteristika najvažniji vid njihova sadržaja. Različite strategije omogućavaju dugoročnu pohranu različitih razina elektroničkih publikacija.

Potrebno je voditi računa i o materijalnoj razini koja se odnosi na skladištenje tijekom podataka. U nastavku su prikazana sva tri spomenuta aspekta dugoročne pohrane.

9.1.2. Skladištenje tijekom podataka

Pri skladištenju mrežne građe valja voditi računa o tom da repozitorij sadrži dovoljan kapacitet memorije i da podržava veću količinu podataka. Njegov sustav mora biti sposoban stvarati sigurnosne kopije podataka, a u slučaju nesreća mora pružiti tehničku podršku za brzo reagiranje. Sustav mora biti sposoban upravljati redundantnom građom (različite verzije ili formati) i automatski provjeravati pogreške. Također je važno da datotekama dajemo nazive i smještamo ih u skladu s arhitekturom sustava.

Što se tiče medija za smještanje podataka, većina web arhiva skladišti se na magnetnim vrpčama. Kod vrpca je nezgodno linearno skladištenje jer pretraživanja mogu dugo trajati. Zato se u novije vrijeme teži smještanju podataka na tvrdim diskovima. Oni podržavaju sve veću količinu podataka, čak do 2,5 TB (naprimjer magnetni disk Seagate, čiji je izlazak na tržište predviđen za 2009. godinu), što znači da se na njima može smjestiti do 60-ak ili više milijuna mrežnih stranica. Također, predviđa se da će novi optički diskovi CD-ROM/DVD imati sve veći kapacitet (naprimjer SVOD HD-DVD do 10TB, koji izlazi na tržište 2008. godine). Životni je ciklus različitih materijalnih nositelja različit, kod magnetnih diskova traje do pet godina, kod

magnetnih vrpca do trideset godina,²⁹⁶ a za optičke diskove, koji se čine najstabilnijima, predviđa se da će im životni ciklus trajati sto godina, iako u praksi zasad nije tako. Stručnjaci u njemačkom IBM-u otkrili su da kod optičkih diskova dolazi do degradacije materijala i zato preporučuju da se podatci migriraju svake dvije godine. Prema Moorovu zakonu²⁹⁷ cijena skladištenja smanjuje se svake dvije godine za 50%, što za nabavu medija ne bi trebalo biti veliki problem.

Životni ciklus različitih materijalnih nositelja ovisi i o temperaturi i relativnoj vlazi mjesta na kojem su uskladišteni. Po Brownu²⁹⁸ optimalni su nositelji CD-ROM-i i DLT vrpce. Na osnovi priručnika *Digital Preservation Coalition*²⁹⁹ na temperaturi od 10 °C i relativnoj vlazi 25 RH dostižu najduži životni ciklus (vidi **tabelu br. 3**), DLT čak do 75 godina. Međutim, prosječna temperatura u skladištima knjižnice je od 15 do 20 °C.³⁰⁰ Na toj se temperaturi već opažaju velike razlike.

Tabela br. 3: Životni ciklus različitih materijalnih nositelja. (Izvor: *Digital Preservation Coalition Handbook: Chapter 5: Media and Formats* URL: <http://www.dpconline.org/graphics/medfor/media.html>)

	25% rel. vlage, 10°C	30% rel. vlage, 15°C	40% rel. vlage, 20°C	50% rel. vlage, 25°C	50% rel. vlage, 28°C
D3 magnetna vrpca	50 godina	25 godina	15 godina	3 godina	1 godinu
DLT magnetna vrpca u kaseti	75 godina	40 godina	15 godina	3 godina	1 godinu
CD/DVD	75 godina	40 godina	20 godina	10 godina	2 godina
CD-ROM	30 godina	15 godina	3 godina	9 mjeseci	3 mjeseci

²⁹⁶ Blau, J. Do Burned CDs Have a Short Life Span?. In: PC World, 10.1.2006. URL: <http://msn.pcworld.com/article/id.124312-page.1/article.html>. (15.12.2006.)

²⁹⁷ Stokes, H. Understanding Moore's Law. URL: <http://arstechnica.com/articles/paedia/cpu/moore.ars/1>. (15.12.2006.)

²⁹⁸ Adrian Brown u smjernicama Digital Preservation Guidance Note 2: Selecting storage media for long-term preservation. – The National Archives (URL: <http://www.nationalarchives.gov.uk/preservation/advice/digital.htm>. (15.12.2006.)) - uspoređuje različite materijalne nositelje iz perspektive trajnosti, kapaciteta, stabilnosti, cijene, osjetljivosti i zastarijevanja.

²⁹⁹ Digital Preservation Coalition Handbook: Chapter 5: Media and Formats <http://www.dpconline.org/graphics/medfor/media.html>. (15.12.2006.)

³⁰⁰ Adrian Brown također daje specifikacije o optimalnim temperaturama i relativnoj vlazi u okviru smjernica *Digital Preservation Guidance Note3: Care, handling and storage of removable media*. – The National Archives URL: <http://www.nationalarchives.gov.uk/preservation/advice/digital.htm>. (15.12.2006.)

Koliko je važno posvetiti pozornost materijalnim nositeljima pokazuje iskustvo NASA-e, koja je zbog dotrajalih magnetskih vrpce izgubila do 20% informacija prikupljenih 1976. godine u okviru misije na Mars Viking. Zbog sličnih su razloga britanski nacionalni arhivi izgubili rezultate trogodišnjeg rada (od 1976. do 1979.) pokušavajući povratiti podatke o popisu stanovništva iz 1960. godine.³⁰¹

9.1.3. Dugoročna pohrana pristupa elektroničkoj građi

Kod dugoročne pohrane pristupa do elektroničke građe nema jedinstvenih rješenja. U svakom slučaju postoji opasnost da se podatci izgube.

S obzirom na prioritete pohrane podataka, metode održavanja pristupa možemo podijeliti u dvije skupine: s jedne strane je strategija čiji je cilj pohraniti logičku razinu, znači sintaksu zapisa elektroničkog dokumenta, a s druge je strategija čiji je prioritet čuvanje sadržaja, znači konceptualne razinu ili razine specifičnih karakteristika dokumenta, bez obzira na logički zapis. U nastavku su prikazane moguće metode u okviru navedenih strategija.

9.1.3.1. Strategija čiji je cilj dugoročna pohrana strukture logičkoga zapisa

Ako je prioritet čuvanje logičke razine, onda se moramo pobrinuti da sačuvamo i tehnološko okruženje u kojem se digitalna građa može čitati ili dodati novoj informacijskoj tehnologiji programske ili strojne dodatke, koji će nam omogućiti da novijom tehnologijom oponašamo djelovanje stare.

Među poznatijim strategijama čiji je cilj sačuvati prvobitni logički zapis jesu: dugoročna pohrana informacijske tehnologije ("muzejski pristup"), tiskanje na papiru, emulacije reinkarnacija starih strojeva i softverski inženjering (VERS, UVC).

Dugoročna pohrana informacijske tehnologije ili "muzejski pristup" je čuvanje okruženja koje je omogućilo pristup određenim digitalnim sadržajima. Problem kod ovog pristupa je taj što se tehnologija mijenja svakih pet godina, a znanje je vezano uz generaciju stručnjaka. Ovaj

³⁰¹ Aschenbrenner, A. Long-term preservation of digital material - building an archive to preserve digital cultural heritage from the Internet. Master Thesis, Technical University Vienna, Dec. 2001. URL: <http://www.ifs.tuwien.ac.at/~aola/publications/thesis-ando/Contents.html>. (15.12.2006.)

pristup postaje veoma skup nakon deset godina, ili prije, jer je teško naći stručnjake i rezervne dijelove za održavanje tih strojeva.

Tiskanje na papiru i danas je često, međutim nije rješenje za elektroničke publikacije koje imaju važne funkcije u digitalnom okruženju.

Emulacije - pomoću dodataka u strojnoj ili programskoj računalnoj opremi pokušavamo postići da se određena tehnologija ponaša kao druga (naprimjer novi operacijski sustavi). Ova je strategija veoma skupa jer su emulatori ili čipovi elementi koje treba sastaviti ili aplikacije koje treba napisati. Za svaku novu platformu ili tehnologiju potreban je nov emulator, što dovodi do slaganja emulatora na emulator ili migracije emulatora, a ne jamči dobre rezultate. Teško je govoriti o emulaciji kao rješenju kod složenih sustava i tehnologija sastavljenih od više platformi, jer je emulacija veoma skupa i nije sasvim uspješna. Svaki nov emulator zahtijeva i dodatno održavanje.

Apstrakcija podataka - Raymond Lorie³⁰² predlaže da se u vrijeme skladištenja elektroničke publikacije napravi program u jeziku univerzalnoga virtualnoga računala (UVC – *Universal Virtual Computer*) koji sadrži opis pravila kodiranja i dekodiranja podataka u formatima u kojima su napisani dokumenti. Strojni bi jezik trebao biti tako jednostavan da ga u budućnosti pomoću UVC interpretera može pročitati bilo koje računalo (vidi **sliku br. 12 i 13**). Nedostatak ovog pristupa je taj što zbog pojednostavljivanja dolazi do očuvanja ograničenoga broja funkcija, a usto su potrebna dodatna sredstva za stvaranje programa u jeziku UVC.

Metode enkapsuliranja - digitalni se objekt smješta u određen format (može biti XML ili koji drugi) s referentnim informacijama, odnosno metapodacima koji sadrže upute i informacije o tom kako mora izgledati original. Daje rješenje za složene baze podataka i multimedijских objekata.

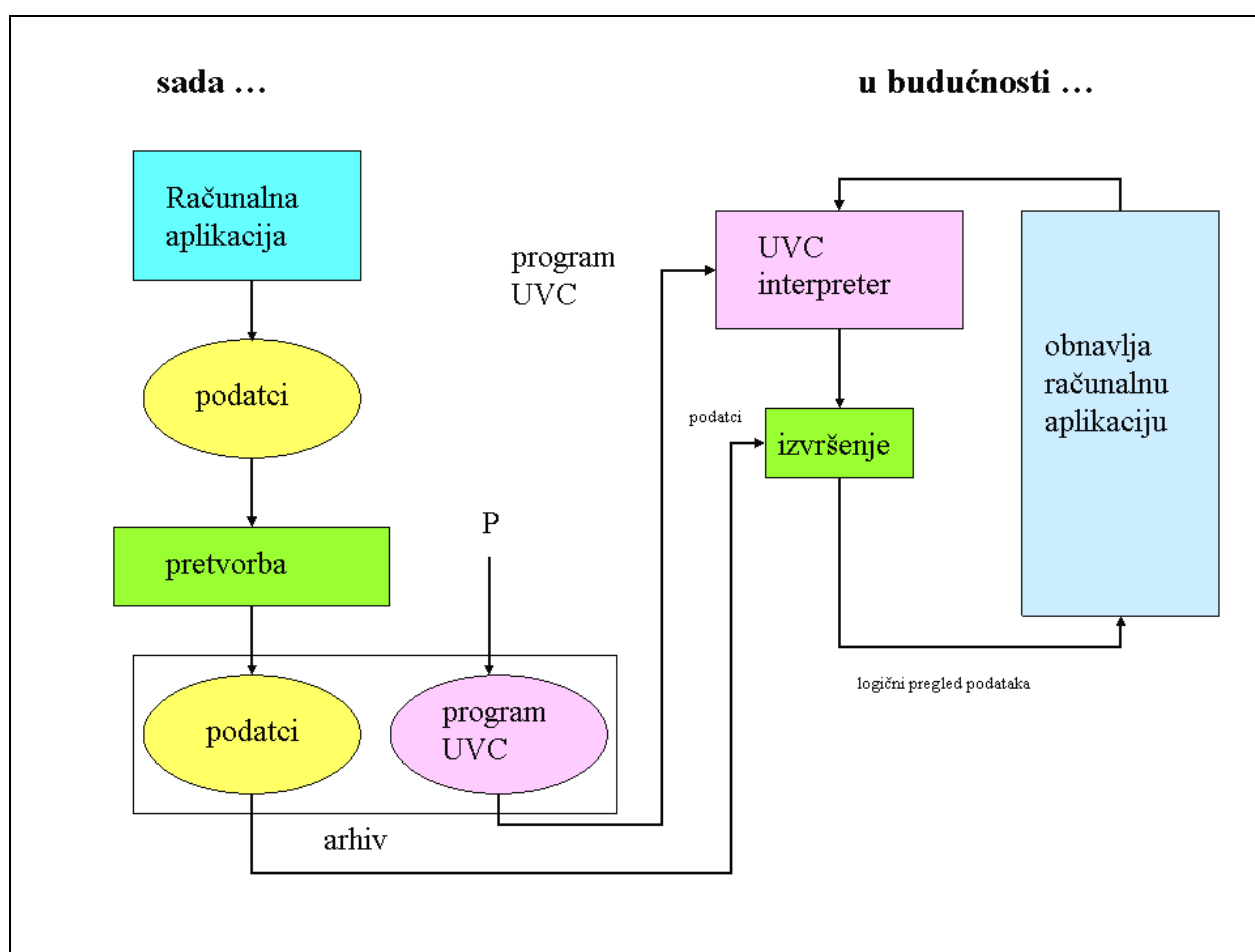
Metoda koja se temelji na načelu prijevoda kamena iz Rosette³⁰³ - razvili su je u *San Diego Supercomputer Center*, u SAD-u, a temelji se na popisu reprezentativnih uzoraka podatkovnih tipova sa svim svojstvima. Uzorci se ponavljaju u drugim formatima koje može pročitati trenutna računalna tehnologija i čine skupinu referenca za potencijalno dekodiranje starijih formata.

³⁰² Lorie, R. A. *A Project on Preservation of Digital Data*. URL: <http://www.rlg.org/preserv/diginews/diginews5-3.html#feature2>. (15.12.2006.)

³⁰³ Metoda koja se temelji na načelu prijevoda kamena iz Rosette dobila je ime po slavnom kamenu, koji je sadržavao tekst napisan na grčkom jeziku, feničanskim alfabedom i hijeroglifima. Grčki i feničanski tekst omogućili su prvi prevod hijeroglifa.

Metoda povratka podataka (*data archeology*) - ne zahtijeva djelovanje u trenutku arhiviranja jer se pretpostavlja da će u budućnosti stručnjaci biti sposobni da preko istraživanja prevedu i pročitaju sadržaje koji djeluju sa starom informacijskom tehnologijom i aplikacijama.

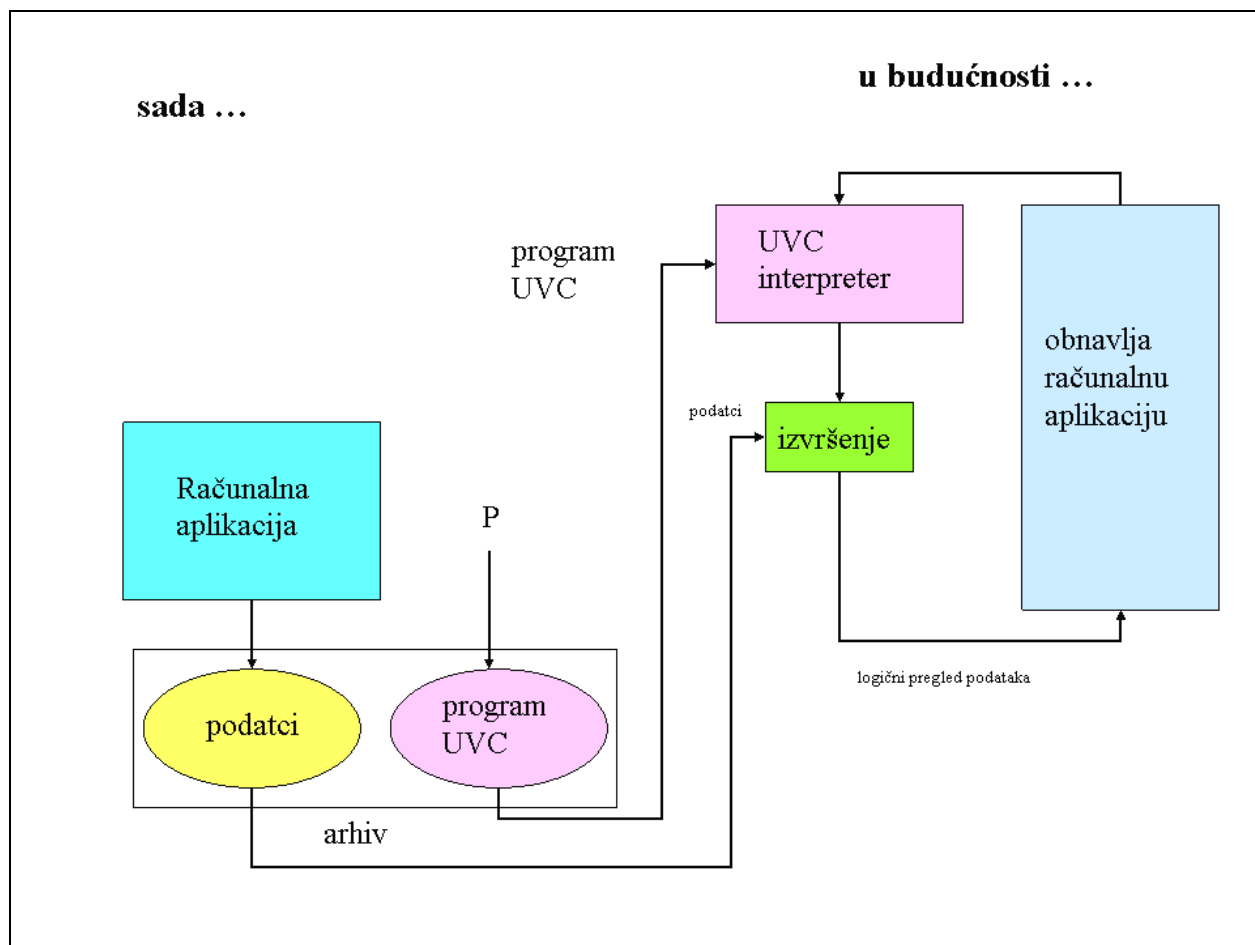
Metode praćenja formata. Već danas postoje različiti popisi formata koji sadrže podatke o informacijsko-tehnološkom okruženju koje je potrebno za njihovo djelovanje, o njihovim nakladnicima ili autorima, o novijim kompatibilnim verzijama i slično. Među njima su baza PRONOM, koju od 2002. godine održavaju britanski nacionalni arhivi³⁰⁴ i baza koju su izgradili u Sveučilišnoj knjižnici u Harvardu, u okviru projekta JHOVE (*JSTOR/Harvard Object Validation Environment*).³⁰⁵



Slika br. 12. Scenarij 1. arhiviranja podataka. (Izvor: Lorie R. A. *A Project on Preservation of Digital Data*. RLG DigiNews. 5, 3(2001). URL: <http://www.rlg.org/preserv/diginews/diginews5-3.html#feature2>, Figure 2)

³⁰⁴ Vidi mrežne strane; <http://www.nationalarchives.gov.uk/PRONOM/Software/proSoftwareSearch.asp?status=new> i <http://www.nationalarchives.gov.uk/pronom/about.htm>. (15.12.2006.)

³⁰⁵ Abrams, S. L. The role of format in digital preservation. // *VINE: The Journal of Information and Knowledge Management Systems*, 34, 2(2004), str. 49-55. Baza je dostupna na URL: <http://hul.harvard.edu/jhove/>. (15.12.2006.)



Slika br. 13. Scenarij 2 arhiviranja podataka. (Izvor: Lorie R. A. *A Project on Preservation of Digital Data*. RLG DigiNews. 5, 3(2001). <http://www.rlg.org/preserv/diginews/diginews5-3.html#feature2>, Figure 3)

9.1.3.2. Strategija čiji je cilj dugoročna pohrana sadržaja zapisa, odnosno konceptualnoga objekta

Kod ovih se strategija u svakom slučaju mijenja logički objekt i pokušava se održati originalni izgled sadržaja.

Migracija je znači prijenos podataka iz jedne generacije softvera, odnosno hardvera, u drugu. Pritom se mijenja logični format, dok je konceptualni objekt predstavljen s novom strojnom ili računalnom opremom. Do migracije dolazi prijelazom u novi format (naprimjer SGML >HTML>XML) ili prijelazom u novu verziju formata (naprimjer Word 95>Word 97> Word 2000). Do migracije dolazi kod standardizacije formata, iako se pritom gube i neke bitne funkcije građe (naprimjer migracija u ASCII ili PDF/A).

Pretvorba na osnovi TOM-a (*Typed Object Model*) - podrazumijeva razvrstavanje formata među kojima je moguća migracija na osnovi njihovih svojstava.

Reinženjering softvera služi za migraciju softvera i stvaranje novih platformi ili operacijskih sustava.

Format za razmjenu objekata - XML (*Extensible Markup Language*) – to je jezik koji je nastao iz SGML-a (ISO 8879) i služi za bogaćenje podataka s informacijama o strukturi i sadržaju. To je standard otvorenoga tipa koji je definirao *World Wide Web Consortium* i neovisan je o platformi ili programu za djelovanje. U nizozemskoj Kraljevskoj knjižnici³⁰⁶ razmotrili su upotrebu XML-a kao formata za dugoročnu pohranu elektroničke građe. Njegove su prednosti te što je veoma prilagodljiv, u njemu je moguće enkapsulirati različite vrste metapodataka o digitalnom objektu i kompatibilan je sa svim oblicima hipertekstualnoga jezika. Ima veoma jasna pravila, njegova je struktura definirana određenim DTD-ima (*Document Type Definition*) u koje se mogu ugraditi metapodatci tipa MARC, a može se čitati kako golim okom tako i pomoću interpretera. Služi i kao format za mapiranje digitalnih objekata, posebice složenijih, multimedijjskih, za izmjenu podataka s drugim sustavima ili za migraciju tih objekata, tako što se bilježi njihova struktura, format i sadržaj.

U okviru projekta *Making of Amerika II* (MOA2) 2001. godine nekoliko američkih sveučilišta³⁰⁷ izradilo je XML DTD shemu za kodiranje deskriptivnih, administrativnih i strukturalnih metapodataka za tekstove i slike.³⁰⁸ U to je vrijeme predmet javne rasprave bio referentni model za otvoreni arhivski informacijski sustav (OAIS), koji je predviđao pakiranje (*packaging*) digitalnih objekata u informacijske pakete (*information packages*) u digitalnom repozitoriju. Poslije, na inicijativu američke *Digital Library Federation*, na osnovi tog su formata izradili shemu METS (*Metadata Encoding & Transmission Standard*).³⁰⁹ METS se sastoji od pet glavnih sekcija:

³⁰⁶ Slat, J. ; Bakker, H. XML and digital preservation: Digital preservation testbed White Paper. Den Haag : Digital Preservation Testbed Project, 2002. URL: http://www.digitaalduurzaamheid.nl/bibliotheek/docs/white-paper_xml-en.pdf. (15.12.2006.)

³⁰⁷ U spomenutom projektu MOA2 su sudjelovali: Sveučilište Berkeley u Kaliforniji, Sveučilište Cornell i Javna knjižnica u New Yorku, Pensylvanijsko Državno Sveučilište i Sveučilište u Stanfordu

³⁰⁸ Cantara, L. METS : The Metadata Encoding and Transmission Standard. // *Cataloging & Classification Quarterly*, 40, 3-4(2005), str. 237-253.

³⁰⁹ Više informacija o METS na mrežnim stranicama: <http://www.loc.gov/standards/mets/>. (15.12.2006.)

- deskriptivni metapodatci - u kojima je moguće kodirati bibliografske metapodatke iz drugih shema, naprimjer MARC, UNIMARC ili EAD;
- administrativni metapodatci - koji sadrže informacije o nadnevku stvaranja i smještanja podataka u arhiv, intelektualnom vlasništvu, originalima;
- skupine datoteka - podatci o formatima datoteka i različitim verzijama;
- strukturalne mape - najvažniji dio koji pokazuje hijerarhijsku strukturu datoteka digitalnoga objekta i njihovih metapodataka;
- ponašanja - informacije o djelovanju digitalnoga sadržaja objekta u arhivu.

METS je nastao kao shema za kodiranje ili pakiranje digitalnih objekata u digitalnom arhivu koji slijedi procese OAIS,³¹⁰ i to s ciljem da olakša izmjenu digitalnih objekata iz jednoga repozitorija u drugi.³¹¹ Prednost je u tome što XML nije vezan za platformu ili program pa je njegova upotreba dobrodošla kod dugoročne pohrane. Usto je prijenos metapodataka iz jednog bibliografskog sistema u drugi s ovom shemom jednostavan. METS je prihvaćen kao standard i održava ga *Network Development and MARC Standards Office*³¹² Kongresne knjižnice. U Kongresnoj su knjižnici prilagodili METS za kodiranje zapisa u MARC 21, dok u portugalskoj Nacionalnoj knjižnici pakiraju digitalne objekte s bibliografskim opisom u UNIMARC-u.

U okviru projekta Kopal³¹³ u njemačkoj su Nacionalnoj knjižnici definirali univerzalni format objekta (UOF) u XML-u koji se temelji na standardu METS, a koristi se metapodacima za dugoročnu pohranu iz sheme LMER.³¹⁴

Jednostavnija verzija sheme METS je MODS (Metadata Object Description Schema)³¹⁵ koja se koristi odabirom elemenata u shemi METS. Iako je MODS veoma pojednostavljena shema,

³¹⁰ Referentni model OAIS operira s konceptom informacijskih paketa, koji kruže kroz cijeli sustav i pritom se mijenjaju. Informacijski su objekti su pakirani s metapodacima i u trenutku ulaska u sustav OAIS ih definira kao SIP (*Submission Information Package*). Smještaju se u repozitorij kao AIP (*Archival Information Package*), dostupni su kao DIP (*Dissemination Information Package*).

³¹¹ Cantara, L., op. cit. str. 239. Zanimljiv je članak Seadle, M. METS and the metadata marketplace. // *Library Hi Tech*, 20, 3(2002), str. 255-257, u kojem se još uvijek ne zna kakva će biti budućnost referentnog modela OAIS i dvoji se o budućnosti METS-a.

³¹² <http://www.loc.gov/marc/ndmso.html>. (15.12.2006.)

³¹³ <https://kopal.langzeitarchivierung.de/index.php/en>. (15.12.2006.)

³¹⁴ Steinke, T. The universal object format - an archiving and exchange format for digital objects. // *ECDL 2006 / Gonzalo, J. ... [et al.] (Eds.)*, 2006. (LNCS 4172). Str 552-554

³¹⁵ <http://www.loc.gov/standards/mods/>. (15.12.2006.)

METS je bogatiji od *Dublin Core* i prijenos podataka u MARC 21 je jednostavan. U obrnutom smjeru, iz MODS-a u MARC 21, kao i u primjeru *Dublin Core*, dolazi do gubitka podataka. Koristi se za različite digitalne objekte, naprimjer elektroničke knjige, serijske publikacije, računalne programe, konferencijske publikacije, kartografsku građu, video i glazbene datoteke, zvučne datoteke i ostalu multimedijску građu. Na stranicama Kongresne knjižnice različite su sheme za prijenos iz MODS-a ili METS-a u druge formate, i obratno.

Upravo je izašla novija verzija MADS-a (*Metadata Authority Description Schema*)³¹⁶ za potrebe normativne kontrole odrednica.

Upotreba XML sheme METS sve je više rasprostranjena ne samo u SAD-u, već i u Europi. U većini slučajeva primjenjuje se kod digitalizirane građe ili građe dobivene neposredno od nakladnika. Također služi i pri ručnom prikupljanju digitalnih objekata s mreže, gdje je moguća veća kontrola pojedinačnih jedinica građe. U trenutku unosa digitalnoga objekta u arhivski sustav najčešće se primjenjuju migracije ili standardizacija formata. Kod pojedinačnih elektroničkih publikacija planiranje strategije dugoročne pohrane mnogo je jednostavnije nego kod mrežnih arhiva, koji su automatski prikupljeni pomoću robota. Kod tih arhiva javlja se problem jer mnoge domaće mrežne stranice ili mrežna mjesta sadrže različite formate ili su vezani na složenu tehnologiju, koju roboti ne mogu prepoznati. Njihovi moduli za indeksiranje u arhivskim sustavima omogućavaju izlučivanje datoteka u određenom formatu. Međutim, opća migracija verzije ili formata mogla bi utjecati na kvalitetu prikazivanja njihova sadržaja. U konzorciju su našli rješenje za ovaj problem tako što su izradili format WARC u XML-u, u koji se automatski smještaju podatci koje prikuplja robot.

Bez obzira na to koju od metoda dugoročne pohrane elektroničke građe odaberemo, moramo posvetiti pažnju sljedećim kriterijima:³¹⁷

- **Izvedivost metode** – mora postojati odgovarajuća programska, odnosno računalna oprema (softver i hardver) za njezino izvođenje;
- **Trajnost metode** – mora se prilagođavati tehnološkim promjenama i drugim metodama;
- **Praktičnost** – njezina složenost i cijena moraju biti primjerni vrijednosti sadržaja;

³¹⁶ <http://www.loc.gov/standards/mads/> . (15.12.2006.)

³¹⁷ Thibodeau, K. L. *Overview of technological approaches to digital preservation and challenges in coming years*. URL: <http://www.clir.org/pubs/reports/pub107/thibodeau.html> . (15.12.2006.)

- **Primjernost** – metoda bora biti primjerena vrsti građe koju želimo dugoročno sačuvati.

Zbog veće sigurnosti, preporuča se, ako je moguće, čuvanje originala svih dokumenata, a konverzije se izvode na njihovim kopijama. Ne znamo u kojem će smjeru tehnologija nastaviti svoj razvoj i hoće li nam biti lakše vratiti se nekoliko godina unatrag te ponovno odabrati pravi put.³¹⁸

9.2. Koncept referentnoga modela OAIS

Na inicijativu Međunarodne organizacije za standardizaciju (ISO) međunarodna udruga svemirskih agencija CCSDS (*Consultative Committee for Space Data Systems*) pripremila je prijedlog referentnoga modela elektroničkoga arhiva u svrhu standardizacije izgradnje digitalnih arhiva za dugoročnu pohranu elektroničke građe. CCSDS je pripremio referentni model koji bi trebao odrediti terminologiju i koncepte za opis i usporedbu modela podataka i arhivskih arhitektura, identificirati veće entitete i njihove međusobne odnose u arhivskom okruženju, pojasniti ključne funkcionalne i informacijske komponente arhivskoga sustava te konačno ponuditi okvir u kojem bi se standardizacija odvijala.

U svibnju 1999. godine CCSDS je dovršio referentni model OAIS (*Open Archival Information System*), koji opisuje okruženje u kojem djeluje arhiv,³¹⁹ funkcionalne komponente arhiva i informacijsku infrastrukturu, koja podupire procese u arhivu. Još u vrijeme osnutka prijedlog OAIS-a bio je predmet rasprave šire stručne javnosti izvan svemirskih organizacija. Osnutak je vrlo brzo privukao pažnju knjižnica i ostalih institucija s kulturnim sadržajima, vladinih agencija i privatnoga sektora.³²⁰ Godine 2001. izašla je nova obnovljena verzija Leta referentnoga modela, a 2003. godine prijedlog je registriran kao ISO standard broj 14721.

³¹⁸ U Nacionalnoj knjižnici Novoga Zelanda čuva se original elektroničke građe bez obzira na to može li se on još pročitati, a s njim i varijantu koja je konverzijom prilagođena novoj tehnologiji.

³¹⁹ Termin "arhiv" u ovom radu koristimo za označavanje sustava za pohranu elektroničke građe.

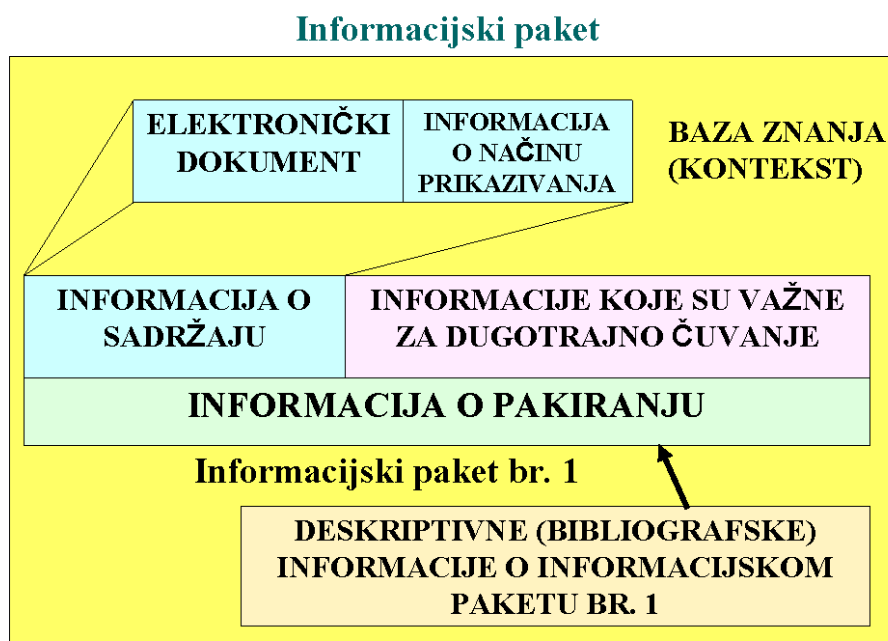
³²⁰ Trusted Digital Repositories: Attributes and responsibilities : an RLG-OCLC Report (2002). Mountain View, CA : RLG. URL: <http://www.rlg.org/longterm/repositories.pdf>. (5.4.2004.), str. 5.

Standard definira OAIS kao "arhiv, koji je sastavljen od organizacije ljudi i sustava i koji je preuzeo odgovornost da dugoročno pohranjuje informacije i omogućava određenoj zajednici pristup do njih".³²¹ Referentni model OAIS teorijski je model koji obuhvaća sve procese u cijelom sustavu za dugoročnu pohranu elektroničke građe.³²²

OAIS se temelji na pohrani informacija zajedno s kontekstom, odnosno bazom znanja koja je potrebna za razumijevanje sadržaja tih informacija. Standard čine informacijski model, koji definira vrste informacija ili informacijskih objekata u sustavu, te funkcionalni model, koji opisuje proces u sustavu.

9.2.1. Informacijski model

Informacijski model OAIS razlikuje četiri vrste informacijskih objekata koji se odnose na različite vrste informacija potrebnih za dugoročnu pohranu elektroničke građe (vidi **sliku br. 14**):



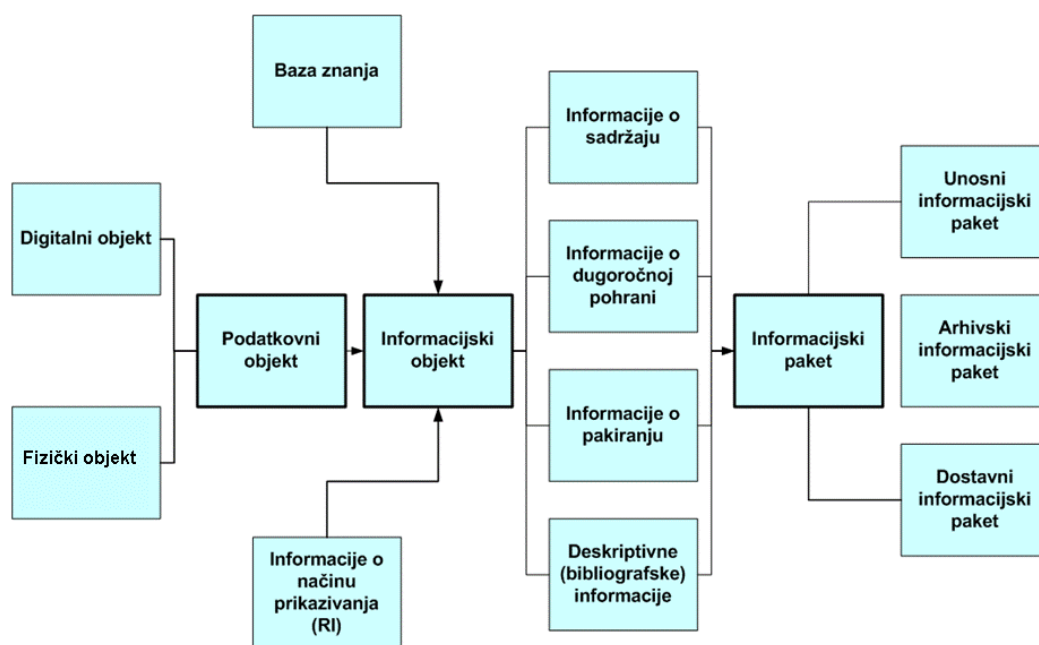
Slika br. 14: Struktura informacijskoga paketa. (Izvor: Reference Model of an OAIS, 2002, Slika br. 2-3, str. 2-5.)

³²¹ Reference Model of an Open Archival Information System (OAIS) : Recommendation for space data system standards : CCSDS 650.0-B-1.Blue Book. Washington: CCSDS (Consultative Committee for Space Data Systems), 2002. URL: http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html. (5.4.2004.) Na str. 1-1 preporuka piše: "An OAIS is an archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community."

³²² Sažeta predstava teoretskog modela autorica ovog rada prikazala je u svome članku: Kavčič-Čolić, A. Teoretični model digitalnih arhivov. // Knjižnica. 48, 4(2004), str. 63-75.

1. **Informacije važne za dugoročnu pohranu sadržaja** (*Preservation description information* - PDI) jesu one koje će nam u budućnosti omogućiti da razumijemo informacije o sadržaju. To su:
 - a. **referentne informacije** - omogućavaju identifikaciju sadržaja elektroničke građe (mogu biti i bibliografski identifikatori, naprimjer ISBN, ISSN i dr.);
 - b. **informacije o kontekstu** - govore o razlozima zašto je određena elektronička građa stvorena i kako je povezana s ostalim sadržajima, odnosno ostalim informacijama izvan informacijskoga paketa (naprimjer iz kojeg originala je slika, je li dokument rezultat digitalizacije, je li slika dio elektroničke knjige, je li članak dio serijske publikacije itd.);
 - c. **informacije o podrijetlu** - dokumentiraju povijest sadržaja elektroničke građe od njezina nastanka i vlasništva do promjena i intervencija na njoj (npr. je li dokument elektronička verzija originala ili je upotrijebljen poseban postupak kompresije digitalne slike);
 - d. **informacije o vjerodostojnosti** - (*fixity information*) dokumentiraju mehanizme za dokazivanje autentičnosti informacija o sadržaju i njegovoj nepromjenljivosti (kontrolni zbroj (*checksum*), digitalni potpis ...), koji jamče da se građa nije promijenila od trenutka njezina preuzimanja.
2. **Informacija o sadržaju** je ona koju želimo sačuvati. To je zapravo elektronički dokument ili, ako se držimo terminologije standarda, digitalni objekt. Obuhvaća sljedeće vrste informacija:
 - a. **informacije o informacijskom objektu** (sam digitalni objekat);
 - b. **informacije o načinu prikazivanja sadržaja** (*Representation information* - RI) (struktura i semantika digitalnoga objekta).
3. **Informacije o pakiranju** opisuju način na koji je smješten elektronički dokument, odnosno na koji su smještene njegove komponente u arhivu.
4. **Deskriptivne informacije** sadrže bibliografski opis koji nam omogućava lokalizaciju elektroničke građe.

Svaki informacijski objekt nastao je u okviru određene **baze znanja**. Naprimjer program u Javi nastao je pomoću poznavanja programskoga jezika Java. Budući da je to znanje vrlo opće, iz tih informacija ne možemo saznati što je sačuvano u određenom informacijskom objektu. Zato su potrebne informacije koje nam pomažu da razumijemo podatke koji su sačuvani u elektroničkom obliku (naprimjer je li broj 3548-4995 međunarodni standardni broj ISSN). Svi informacijski objekti sastavljeni su od **podatkovnoga ili digitalnoga objekta** (popis bitova u obliku niza brojeva) i **informacije o načinu prikazivanja** (*Representation information-RI*), koje omogućavaju interpretaciju i razumijevanje sadržaja digitalnoga objekta (naprimjer je li elektronički zapis slika, tekst ili zvučna snimka). Informacije o načinu prikazivanja mogu biti strukturne ili semantičke. Strukturne govore o formatu podataka i mogu opisivati strojno i programsko okruženje koje je potrebno za pristup podacima (npr. digitalni objekt je tekst napisan u ASCII). Semantičke informacije daju smisao podatkovnim strukturama, koje određuju strukturne informacije (naprimjer, tekst je na engleskome jeziku i daje objašnjenje kako moramo interpretirati informacije o pakiranju itd.) (**slika br. 15**).



Slika br. 15: Informacijski model OAIS. (Izvor: Lavoie, B. *Meeting the challenges of digital preservation: The OAIS Reference Model* // OCLC Newsletter. January/February 2000, nr. 243, str. 28.
<http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000001747&reqid=110130>)

Informacijski model općenito određuje tipove informacija ili metapodataka koji su potrebni za osiguravanje dugoročnoga pristupa do skladištenih informacija u elektroničkom arhivu. Model se ne spušta u specifikaciju tehnologije, praktičnih standarda i metapodataka. To je prepušteno samim arhivima s obzirom na njihove sadržaje i funkciju koju obavljaju u svome okruženju.

Osnovni koncept modela OAIS je informacijski paket koji sadrži sva četiri tipa informacijskoga objekta (slika br. 14). Sastoji se od **informacija o sadržaju i informacija koje su važne za dugoročnu pohranu elektroničke građe** (PDI). Oba su dijela povezana u cjelinu i »upakirana« u informacijski paket. Prate ih **informacije o pakiranju** koje navode način pakiranja (upotrijebljena aplikacija ili drugo). Informacija o sadržaju najvažniji je element informacijskoga paketa i cilj je dugoročne pohrane. Sadrži elektroničku građu³²³ (digitalni objekt) koju želimo sačuvati te informaciju o načinu prikazivanja.

9.2.2. Funkcionalni model

Informacijski se paket od unosa u arhivski sustav do prikazivanja na zahtjev korisnika mijenja i preuzima tri različita oblika koji ovise o arhivskom procesu u kojem se nalazi. Ti su procesi prikazani sljedećim funkcionalnim entitetama:

1. **Ulaz u arhivski sustav** jest ulazni entitet u kojem se vrši unos elektroničke građe u arhivski sustav. Sustav preuzima elektronički izvor od proizvođača, autora, posrednika ili na koji drugi način. U to ulazi i automatsko, odnosno ručno preuzimanje mrežne građe pomoću robota prikupljača. Elektroničku građu mogu pratiti i osnovne informacije o njoj. Standard određuje taj skup ulaznih podataka (digitalni objekt i prateće informacije) kao ulazni informacijski paket (*Submission information package* - SIP). SIP može sadržavati sve ostale informacijske objekte ili samo informaciju o sadržaju. Nakon provjeravanja postoji li već u arhivu njegova kopija ili starija verzija, provjerava se njegova kvaliteta, cjelovitost i vjerodostojnost i dodaju mu se metapodatci koji su bitni za dugoročnu pohranu. Iz jednog SIP-a može nastati jedan arhivski informacijski paket ili više njih (AIP), što ovisi o tomu je li publikacija sastavljena iz

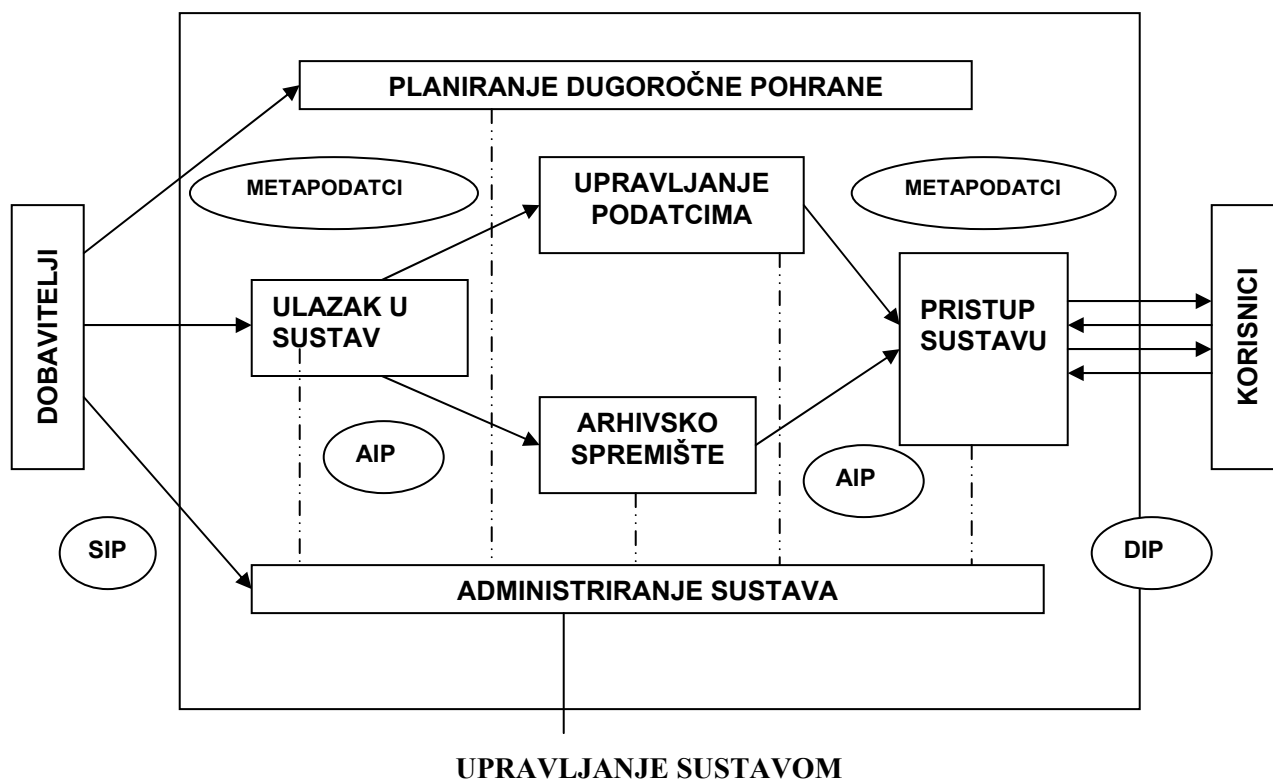
³²³ OAIS kao teorijski model obrađuje elektroničku građu, iako ne isključuje i analognu građu, posebice onu koja nosi u sebi dodatne informacije o elektroničkoj građi (naprimjer priručnici za upotrebu računalnih aplikacija, tiskana slika sučelja, podatkovnih baza, umjetničkih djela itd.). Sustav je tako široko koncipiran da se može prilagoditi svakoj građi.

više samostalnih dijelova. AIP nastavlja svoj put do arhivskoga skladišta ili repozitorija. U ovom se entitetu osim AIP-a spremaju i deksriptivni metapodatci (bibliografski opis), koji se čuvaju odvojeno u posebnoj bazi podataka (naprimjer u katalogu);

2. **Arhivsko spremište (digitalni repozitorij)** preuzima AIP ulazne entitete i skladišti ga na osnovi postojeće hijerarhije skladištenja. Spremište vodi brigu o migracijama medija, rutinskom provjeravanju pogrešaka, omogućavanju povratka podataka u slučaju nesreća i omogućavanju dostupa do AIP-a na zahtjev sustava.
3. **Upravljanje podacima** omogućava pristup deskriptivnim metapodacima (katalogu) fonda i administrativnim podacima koji su važni za upravljanje arhivom. Tu se izvode sve funkcije baza podataka (indeksiranje, pretraživanje, listanje, prikazivanje rezultata i izvještaja ...) i njihovo održavanje, odnosno obnavljanje.
4. **U administriranje** se izvode usluge i funkcije u vezi s djelovanjem cijeloga arhivskoga sustava. Tu se vode zahtjevi i pregovori u vezi s građom, autorima i posrednicima, nakladnicima, provjerava se kvaliteta građe, održava se, odnosno mijenja konfiguracija strojne i programske opreme sustava. U ovom entitetu izvodi se kontrola cijeloga arhiva od preuzimanja ili prikupljanja građe, do skladištenja i njezina pretraživanja i korisničke upotrebe. Usto se odlučuje o obnavljanju ili migraciji u nove verzije ili platforme, odnosno o uvođenju novih standarda. Ukratko, u tom se entitetu definira politika vođenja cijeloga arhivskoga sustava.
5. **Planiranje dugoročne pohrane** osigurava dugoročan pristup korisnika informacijama ili elektroničkim sadržajima koji su uskladišteni u arhivu, bez obzira na to mijenja li se izvorno informacijsko tehnološko okruženje ili ne. To znači da se u tom dijelu mijenja elektronički fond, potrebna infrastruktura za pristup do njega i potrebe zajednice te se određuju standardi i strategije za dugoročnu pohranu elektroničke građe. U slučaju razvoja novih računalnih platformi ili verzija aplikacija, odlučuje se o strategijama, metodama i detaljnim postupcima dugoročne pohrane. Informacije koje nastaju u tim procesima koriste se u ulaznom entitetu radi praćenja i preuzimanja novoga fonda.
6. **Pristup** omogućava korisnicima pretraživanje po katalogu, lokalizaciju građe i pristup građi ili, u slučaju autorski zaštićene građe, ograničen pristup. Korisnik na osnovi bibliografskoga opisa pozove AIP iz arhivskoga skladišta i ako nema ograničenja pristupa može pregledavati cijeli dokument. Pritom AIP promijeni oblik tako da bude primjeren za pregledavanje na

računalnom zaslonu. To je informacijski paket za pristup (*Dissemination information package* - DIP).

Na **slici br. 16** prikazan je model kao što ga navodi standard.



Slika br. 16: Funkcionalni entiteti OAIS (Izvor: Reference Model..., 2002, Slika št. 4-1, str. 4-1)

Digitalnim depozitarnim ustanovama nužno je potreban standard koji bi im ponudio smjernice za proces dugoročne pohrane elektroničke građe. Dokaz za to je da i prije nego što je OAIS prihvaćen kao ISO standard, referentni je model bio korišten u mnogim projektima koji su potvrdili njegovu primjernost pri izgradnji elektroničkih arhiva. Među njima su najpoznatiji CEDARS (*CURL Exemplars in Digital Archives*)³²⁴ iz 1998.-2002. godine, australski projekt PANDORA i europski projekt NEDLIB (1998.-2000.). OAIS je univerzalan model, ne traži određenu infrastrukturu, već se koncentrira na funkcije i procese koji su prihvatljivi za dugoročnu

³²⁴ Day, M. CEDARS: Digital preservation and metadata. // Sixth DELOS Workshop: Preservation of Digital Information, Tomar, Portugal, 17-19 June 1998. URL: <http://www.ukoln.ac.uk/metadata/presentations/delos6/cedars.html>. (5.4.2004.)

pohranu građe i informacije o toj građi na svim medijima. Sve više knjižnica, koje su zadužene za pohranu pisane elektroničke baštine, građe svoje arhive na osnovi referentnog modela OAIS. Kao što smo već vidjeli, nekoliko američkih istraživačkih knjižnica (RLG) i konzorcij OCLC preporučuju ga kao polazište pri razvoju digitalnih arhiva.

Partneri projekta NEDLIB (*NEtworked Digital LIBrary*)³²⁵ u većini su nacionalnih knjižnica primijenili referentni model OAIS kao temelj teorijskoga modela svojih arhiva, s tim što su ga dogradili i prilagodili procesima u knjižnicama. Iskustva na projektu vrlo su opširno opisali u izvještajima koji mogu poslužiti kao smjernice ostalim knjižnicama. Partneri projekta NEDLIB dodali su entitetama iz OAIS-a procese koji su karakteristični za bibliografsku obradu, odnosno za korisnički pristup građi. Tako je nastao DSEP (*Deposit System for Electronic Publications*),³²⁶ depozitarni sustav za elektroničke publikacije, koji je sadržavao jedanaest funkcionalnih entiteta.³²⁷

1. **odabir građe** – to je proces koji se izvodi ručno i na osnovi nacionalne politike dugoročne pohrane građe, odnosno postojeće regulative ili ugovora s vlasnicima autorskih prava;
2. **dobivanje građe** – obuhvaća sve administrativne aktivnosti koje su potrebne za dobivanje elektroničke građe, uključujući dobivanje informacija o novim publikacijama, potpisivanje ugovora, vođenje računovodstva itd.
3. **preuzimanje odnosno prikupljanje na mreži** – predviđa sve aktivnosti koje su potrebne da bi se dobila kopija elektroničke građe. To je preuzimanje građe na materijalnom nositelju (CD, disketa), preko interneta (FTP, e-pošta) ili pomoću robota neposredno iz svjetske mreže, selektivno ili na osnovi domenskoga kriterija;
4. **registracija** – elektronička publikacija unosi se u depozitarni sustav; slijedi registracija, odlučivanje o preuzimanju i obavješćavanje dobavljača o preuzimanju publikacije;

³²⁵ NEDLIB (Networked European Deposit Library) [informacije o projektu]. Pridobljeno 5.4.2004 s spletne strani <http://www.kb.nl/coop/nedlib/>. (5.4.2004.)

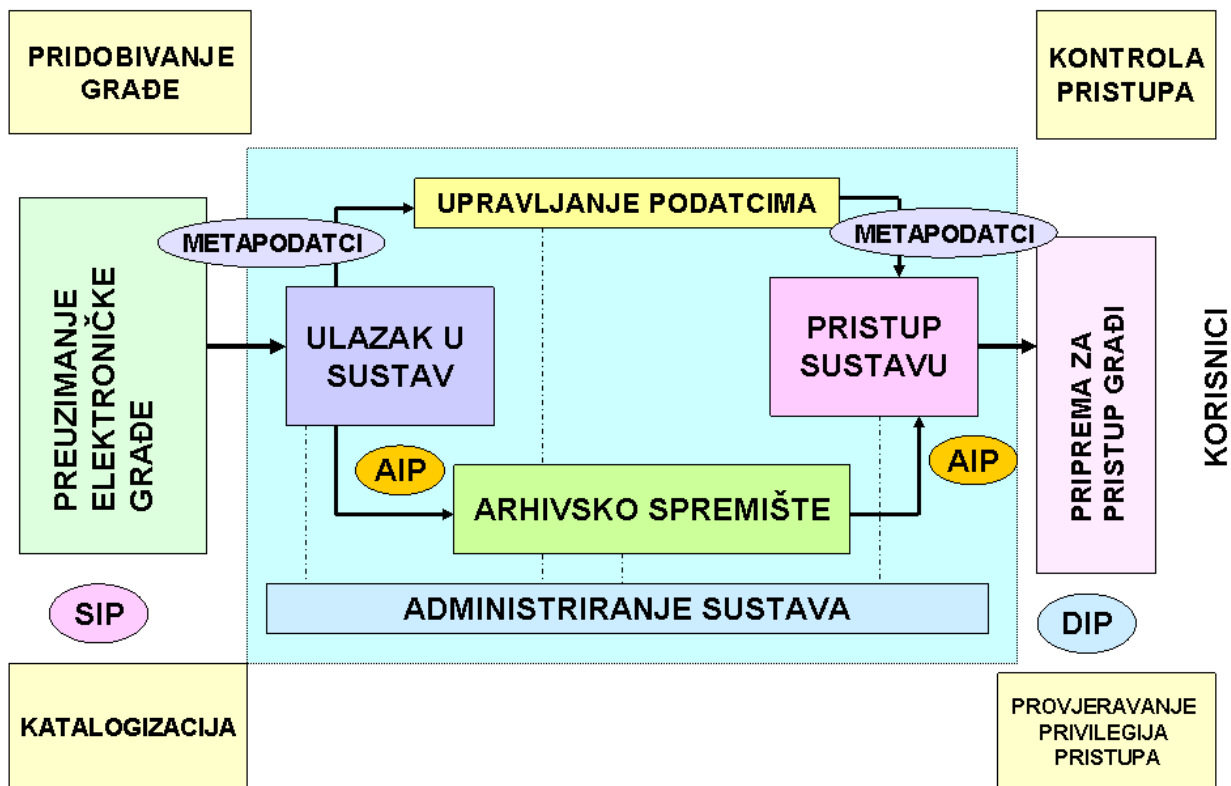
³²⁶ Werf, T. van der. The Deposit System for Electronic Publications: A process Model. Den Haag: Koninklijke Bibliotheek, 2000. (NEDLIB Report Series ; 6). URL: <http://www.kb.nl/coop/nedlib/results/DSEPprocessmodel.pdf>. (5.4.2004.)

³²⁷ Steenbakkens, J. The Nedlib Guidelines : Setting up a Deposit System for Electronic Publications. Den Haag: Koninklijke Bibliotheek, 2000. (NEDLIB Report series ; 5). URL: <http://www.kb.nl/coop/nedlib/results/NEDLIBguidelines.pdf>. (5.4.2004.)

5. **provjeravanje** – to je rutinska kontrola za provjeravanje cjelovitosti i kvalitete dobivene elektroničke građe; ovaj entitet uključuje i procese instalacije, otkrivanje grešaka u djelovanju datoteka i obavještanje nabavnoga odjela;
6. **opis građe** – elektronička se građa katalogizira, pritom se mogu upotrijebiti osnovne informacije koje su dobivene od dobavljača; taj se proces može izvršiti automatski ili poluautomatski, što je slučaj kod prikupljanja mrežne građe robotom. Zatim se unose metapodatci za dugoročnu pohranu, dodaju se identifikatori (URN ili druge), deskriptori itd.;
7. **skladištenje građe** – elektronička se građa šalje u arhivski sustav; u tom se procesu može izvršiti migracija datoteka na drugi materijalni nositelj, kao i njegova reprodukcija i prijenos na drugu lokaciju radi zaštite od nesreća; s vremena na vrijeme kontrolira se njihova kvaliteta;
8. **dugoročna pohrana elektroničke građe** – obuhvaća sve aktivnosti koje su povezane s dugoročnom pohranom, kao što su: planiranje strategije dugoročne pohrane, migracije, utvrđivanje potrebe za emulaciju aplikacija ili konverziju zastarjelih formata, kontrola kvalitete i dr.;
9. **priprema za pristup građi** (*delivery*) – u okviru ovoga entiteta vrši se prijenos kopije originalne građe iz skladišta u međuprostor na računalu, gdje se priprema tako da joj korisnik može pristupiti; ako je, naprimjer, to djelo sastavljeno od više poglavlja, sadrži dijelove koje treba ponovo formatirati ili im možemo pristupiti samo emulacijom, onda se svi ti procesi obave u ovom entitetu;
10. **pristup** – obuhvaća sve funkcije korisničkoga sučelja kojima korisnik potražuje, predaje zahtjev za pristup građi i pristupa joj; poseban je dio kontrola pristupa i provjeravanje privilegija pristupa u slučaju da je građa autorski zaštićena;
11. **upravljanje sustavom** – predviđa upravljanje, kontrolu i praćenje svih deset navedenih procesa.

Na **slici br. 17** prikazan je model DSEP. U središtu ostaje struktura OAIS. Ispred entiteta "ulazak u sustav" i entiteta "pristup" dodani su novi entiteti.

D S E P (Depozitni sustav za elektroničke publikacije)



Slika br. 17: Depozitarni sustav za elektroničke publikacije (Izvor: Werf, 2000, Slika br. 6, str. 17)

Do danas se najviše pažnje posvećivalo metapodacima za dugoročnu pohranu. Teško je predvidjeti promjene informacijske tehnologije i znati koji će nam podatci u budućnosti pomoći u razumijevanju informacija koje danas želimo sačuvati. Zanimljivo je da se u raznim nacionalnim knjižnicama primjenjuju različiti pristupi pri definiciji metapodataka za dugoročnu pohranu elektroničke građe.³²⁸

³²⁸ Dok su metapodatci za dugoročnu pohranu elektroničke građe u okviru projekta NEDLIB određeni na osnovi komponenata sustava OAIS, u australskoj Nacionalnoj knjižnici određuju se na osnovi informacija koje su dobivene izvan sustava. Australci upotrebljavaju 24 vrste metapodataka koji se odnose na različite razine objekta (datoteka, elektronički objekt ili zbirka). Treći primjer je model novozelandske Nacionalne knjižnice koji se temelji na četiri međusobno povezana entiteta, koji određuju važnost metapodataka za dugoročnu pohranu: objekt, proces, datoteka i promjene metapodataka.

9.3. Prilagođavanje repozitorija web arhiva modelu OAIS

Kao što smo vidjeti, teorijski model OAIS namijenjen je izvođenju svih procesa u digitalnom arhivu, uključujući i dugoročnu pohranu elektroničke mrežne građe.

Prije planiranja web arhiva potrebno je utvrditi na koji način želimo smjestiti publikacije koje smo prikupili na mreži. Te se publikacije skladište u obliku arhivskoga paketa (AIP), a pritom treba paziti da svaki AIP čini sadržajnu cjelinu te da je identificiran trajnim identifikatorom, a po mogućnosti i brojem nacionalne bibliografije (NBN). Po Verhoevnu³²⁹ jedan AIP može sadržavati cijelo prikupljanje mreže, domenu ili mrežno mjesto, odnosno URI. OAIS pritom razlikuje zbirku (AIC - *Archival Information Collection*) i jedinicu građe (AIU - *Archival Information Unit*). Kod veće granularnosti može nastati problem ako se izgube veze među stranicama, odnosno veze s okruženjem mrežnoga mjesta. Kod zbirke (AIC-a) teško je razlikovati jedinice i zbog velikog broja podataka upravljanje njima može biti dugoročno i mogu se pritom pojaviti greške. Optimalno je rješenje dodjeljivanje AIP-a ili NBN-a na razini domene jer osiguravaju povezanost URI-a unutarnjih poveznica, unutarnje i vanjske građe. Pritom se građa koja je sastavni dio više mrežnih mjesta ne udvostručuje i vrijeme učitavanja prilikom pristupa je prihvatljivo.

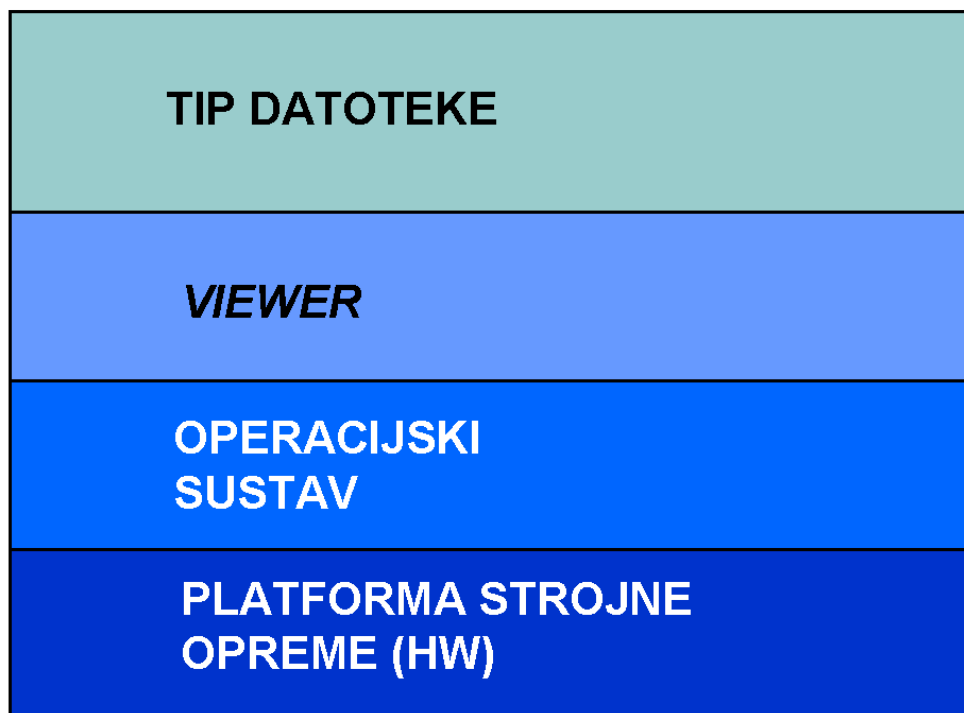
Varijante istoga dokumenta –Novi bi AIP morao upućivati na originalni AIP.

U slučaju različitih formata jednoga dokumenta preporuča se da se svakom formatu dodijeli nov URI. U suprotnom bi nas isti URI vratio na originalni format koji traži posebnu aplikaciju za pristup. Promjena se može izvesti prije arhiviranja ili u trenutku pristupa građi. Ako to napravimo prije arhiviranja, onda moramo AIP pretvoriti u drugi format trajnije vrijednosti Verhoeven predlaže da se promjena obavi u trenutku pristupa. To je dodatni proces, ali on osigurava veću fleksibilnost.

³²⁹ Verhoeven, I. H. Archiving web publications. Amsterdam : IBM ; The Hague : Koninklijke Bibliotheek, 2002. (IBM/KB Long-Term Preservation Study Report Series ; 6). URL: http://www.kb.nl/hrd/dd/dd_onderzoek/reports/6-webpublications.pdf. (15.12.2006.)

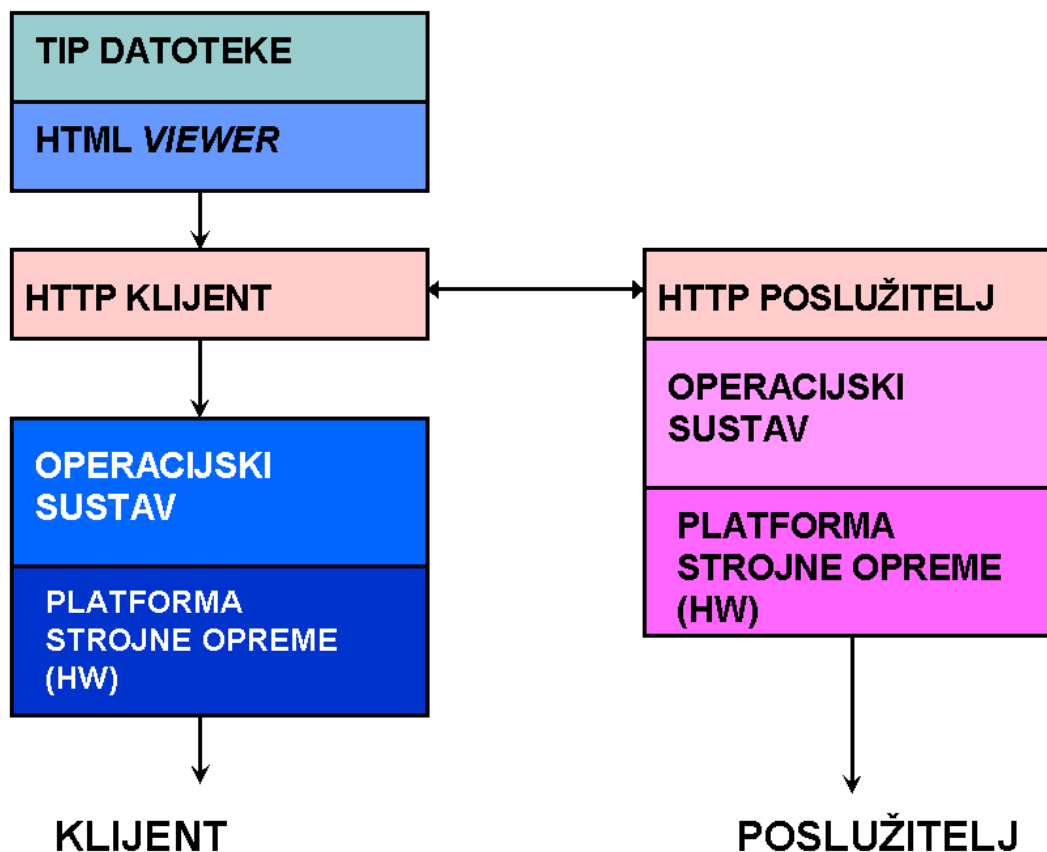
Model slojeva dugoročne pohrane mreže

Kod digitalnih dokumenata treba uzeti u obzir različite slojeve o kojima ovisi pristup njihovim sadržajima. Opći model prikazan je na **slici br. 18**:



Slika br. 18: Primjer modela slojeva dugoročne pohrane elektroničke građe. (Izvor: Verhoeven, Ir. Hans: Archiving web publications. - Amsterdam : IBM ; The Hague : Koninklijke Bibliotheek, 2002. - str. 33. Figure 7.1 / Sample Preservation Layer Model.)

Kada pristupimo digitalnom sadržaju, očitamo ga pomoću interpretera ili viewera. On se oslanja na određen operacijski sustav koji djeluje na odgovarajućoj platformi strojne opreme. Ako uzmemo u obzir da na mreži postoje tisuće formata i nije nužno da su svi čitljivi sa standardnim viewerima moguće je da ćemo ih s vremenom prestati koristiti ili zbog potrebe velikog prostora nećemo održavati zajedno sa standardnim mrežnim zbirkama. U tom su slučaju mogući različiti scenariji. Naprimjer, moguće je da pomoću klijenta pošaljemo ukaz na drugi poslužitelj te da on na zahtjev pokrene potreban viewer i tako nam omogućiti pristup odabranoj datoteci. Znači, viewer se može nalaziti na računalu klijenta, ili na računalu poslužitelja, ili na oba računala, samostalno ili u međusobnoj ovisnosti. (**slika br. 19**)



Slika br. 19: Osnovni model slojeva za dugoročnu pohranu. (Izvor: Figure 7.2 /Basic PLM [Preservation Layer Models] for default HTML file types. In: Verhoeven, Ir. Hans: Archiving web publications. - Amsterdam : IBM ; The Hague : Koninklijke Bibliotheek, 2002. - str.34, Figure 7.2 /Basic PLM [Preservation Layer Models] for default HTML file types.)

Christensen³³⁰ predlaže stvaranje "repozitorija formata" koji bi sadržavali aplikacije za procesiranje različitih formata, viewera i konvertora. U svome članku spominje da je svaka datoteka napisana u određenom programskom jeziku ili formatu i u pozadini ih prate vieweri koji se obično nalaze na poslužiteljima za pristup grafi. Korisnik bi na zahtjev za pristup određenoj datoteci aktivirao odgovarajući viewer ili konvertor sa odgovarajućim viewerom za pretvoreni format i tako bi vrlo brzo došao do željenih podataka. Naravno, s vremenom bi broj formata, a time i viewera i konvertora, znatno porastao. Različite kombinacije viewera i konvertora Christensen ilustrira pomoću dijagrama koji zbog njegova oblika naziva dijagram "T". S jednim viewerom ili konvertorom možemo pristupiti različitim formatima. Zato je nužno potrebno u

³³⁰ Christensen, N. H. Towards format repositories for web archives. // 4th IAWW 2004. URL: <http://www.iwaw.net/04/Christensen.pdf>. (15.12.2006.)

svakom digitalnom repozitoriju istovremeno pripremiti bazu ili tabelu s kombinacijama viewera i formata koji se mogu upotrijebiti za pristup digitalnom sadržaju, kao i tabelu kombinacija konvertora te ulaznih i izlaznih formata odnosno viewera. Christensen usto predlaže da bi bilo korisno imati neku vrijednosnu tabelu s informacijama o uspješnosti migracije, možda čak i s postotcima. Idealno bi bilo kad bi proces odlučivanja i odabira najboljeg viewera ili konvertora bio dio automatskih računalnih procesa pri ulasku do digitalnih sadržaja. Kao što je već spomenuto, postoje javne baze podataka s opisima svih postojećih formata datoteka. Među najpoznatijima je PRONOM u nacionalnim arhivima u Velikoj Britaniji.

10. METAPODATCI I TRAJNI IDENTIFIKATORI

Termin "metapodatak" uzet je iz računalne terminologije i znači doslovno "podatak o podacima". Anne J. Guilliland-Swetland³³¹ definira metapodatke kao "zbir svega što možemo reći o bilo kojem informacijskom objektu³³² na bilo kojoj razini agregacije". Termin se počinje rabiti s nastankom i sve većom upotrebom i širenjem elektroničke građe. Metapodatci pomažu pri identifikaciji, opisu i lokalizaciji građe. Osim toga, omogućavaju nam kontrolu nad elektroničkom građom tako što nam daju informacije o njezinoj strukturi, djelovanju, okruženjima i uvjetima pristupa. Usto određuju tko joj i kako smije pristupiti. Bibliotekari se već godinama koriste metapodacima u bibliografskom opisu građe.

Na osnovi referentnoga sustava OAIS potrebni su nam različiti tipovi metapodataka za praćenje digitalnih objekata u arhivskom sustavu. Općepoznata podjela po funkciji dijeli metapodatke na:

- **deskriptivne metapodatke** - pomažu nam kod opisa i identifikaciji građe u repozitoriju; danas se koristi više shema koje omogućavaju opis građe, kao što su COMARC, Dublin Core ili EAD;
- **strukturne metapodatke** - proslijeđuju informacije o strukturi elektroničke građe; te su informacije vrlo važne kod složenih oblika elektroničke građe, naprimjer u elektroničkim knjigama, zbornicima ili multimedijskoj digitalnoj građi;
- **administrativne metapodatke** - koriste se za upravljanje i administriranje elektroničkom građom; proslijeđuju informacije o nabavi, lokaciji, verzijama i slično;
- **tehničke metapodatke** - proslijeđuju informacije o privilegiju pristupa građi, formatima, djelovanju sustava itd.

S razvojem većih nacionalnih arhivskih sustava i njihovih repozitorija krajem devedesetih godina sve se veća pažnja posvećuje metapodacima, čiji je primarni cilj riješiti problem nečitljivosti elektroničke građe u budućnosti. Taj zadatak imaju **metapodatci za dugoročnu pohranu** (*long-term preservation metadata*), koji su neodvojiv dio svakoga većega elektroničkoga arhiva, koji

³³¹ Guilliland-Swetland, A. J. *Introduction to Metadata: Setting the Stage*, 2000. (Pathways to Digital Information). URL: <http://www.getty.edu/research/institute/standards/intrometadata>. (5. 5. 2004.)

³³² Informacijski objekt je samostalna jedinica građe koja je sastavljena iz jednog ili više elemenata (datoteka) i kojom možemo upravljati pomoću čovjeka ili sustava. Svi informacijski objekti imaju tri karakteristika: sadržaj, kontekst i strukturu.

ima namjeru preživjeti barem pedeset godina, ako ne i više. Za razliku od nabrojenih metapodataka, metapodatci za dugoročnu pohranu sadrže informacije o formatu, odnosno o strukturi bitnoga toka (*bit stream*) digitalnoga objekta, uvjetima ili karakteristikama aplikacije kojom možemo otvoriti elektronički dokument ili o posebnostima i funkcionalnostima elektroničke građe, njezinim izmjenama i slično.

Metapodatci za dugoročnu pohranu elektroničke građe obično su sastavni dio arhivskoga informacijskoga paketa koji stvaramo u trenutku kada arhiviramo elektroničku građu. Također se mogu smjestiti odvojeno od digitalnoga objekta. Mnogi se od metapodataka generiraju automatski u trenutku skladištenja u arhivski sustav. Neke aplikacije ili roboti za prikupljanje elektroničke građe mogu se posebno prilagoditi tako da s unosom generiraju određene podatke.

U zadnjih nekoliko godina pokušao se pronaći optimalan izbor metapodataka koji bi sadržavao informacije koje su važne za pristup elektroničkoj građi uz pomoć novijih informacijskih tehnologija. Projekti čiji je cilj bio rješavanje problema dugoročne pohrane elektroničke građe (CEDARS, NEDLIB i drugi) veliku su pažnju posvetili tim metapodacima. Problematika metapodataka za dugoročnu pohranu elektroničke građe česta je tema u raspravama na međunarodnim skupovima, koji se bave područjem izgradnje digitalnih knjižnica (naprimjer ECDL - *European Conference on Research and Advanced Technology for Digital Libraries* i oELAG - *European Library Automation Group*). U okviru različitih projekata, udruženja ili nacionalnih knjižnica nastale su različite sheme koje potječu iz informacijskoga modela OAIS. Međutim, mnoge definicije te vrste metapodataka polaze od različitih interpretacija entiteta referentnoga modela OAIS.

Projekt CEDARS (*CURL (The Consortium of University Research Libraries) Exemplars in Digital ArchiveS*) (travanj 1998. - ožujak 2002.) jedan je od prvih projekata koji je istraživao strategije za dugoročnu pohranu elektroničke građe. Glavni ciljevi projekta bili su:³³³

- upozoriti na važnost dugoročne pohrane elektroničke građe;

³³³ Cedars Guide to preservation metadata, 2002. URL: <http://www.leeds.ac.uk/cedars/guideto/metadata>. (15.12.2006.)

- identificirati, dokumentirati i proširiti strateški okvir za razvoj politike upravljanja, čiji je cilj dugoročna pohrana elektroničke građe;
- istražiti, dokumentirati i promicati metode za dugoročnu pohranu različitih vrsta elektroničke građe.

Stručnjaci koji su sudjelovali u projektu CEDARS tvrdili su da odluke o dugoročnoj pohrani ne smiju biti uvjetovane metapodacima. Postoje dokumenti koji su vrlo dragocjeni iako nemamo nikakvih informacija o njihovom autoru ni o naslovu.

Projekt CEDARS također se temelji na referentnom modelu OAIS, koristi se istom terminologijom kod definicije informacijskih objekata. Informacije o sadržaju upakirane su zajedno s referentnim informacijama, kao i kod OAIS-a.

Elementi metapodataka određeni su na razini informacijskog paketa³³⁴ (vidi **tabelu br. 4**).

³³⁴ Metadata for digital preservation: the CEDARS project outline specification. Draft for public consultation / The Cedars Project Team and UKOLN, 2000. URL: <http://www.leeds.ac.uk/cedars/MD-STR~5.pdf>. (15.12.2006.)

Tabela br. 4: Shema metapodataka za dugoročnu pohranu elektroničke građe po CEDARS-u

VRSTE INFORMACIJA	METAPODATCI
INFORMACIJSKI PAKET	
1. INFORMACIJE KOJE SU VAŽNE ZA DUGOROČNU POHRANU ELEKTRONIČKE GRAĐE (PDI):	
1.1 REFERENTNE INFORMACIJE:	1.1.1. Opis građe (katalogizacija u postojećem sustavu) 1.1.2. Postojeći metapodatci (bili su već kreirani u katalogu ili na drugom mjestu) 1.1.2.1. Postojeći zapisi o građi
1.2. INFORMACIJE O KONTEKSTU:	1.2.1. Povezanost s ostalim informacijskim objektima
1.3. INFORMACIJE O PODRIJETLU:	1.3.1. Povijest nastanka: 1.3.1.1. Razlozi nastanka 1.3.1.2. Povijest skladištenja 1.3.1.3. Povijest promjena prije arhiviranja 1.3.1.4. Originalno tehničko okruženje 1.3.1.4.1. Uvjeti za djelovanje (hardver/softver) 1.3.1.4.2. Postupci (instalacija opreme ili drugo) 1.3.1.4.3. Dokumentacija (priručnici ...) 1.3.1.5. Razlozi za dugoročnu pohranu 1.3.2. Povijest upravljanja i rukovanja: 1.3.2.1. Povijest procesa unošenja u sustav 1.3.2.2. Povijest administriranja 1.3.2.2.1. Povijest aktivnosti 1.3.2.2.2. Povijest politike koja utječe na odlučivanje 1.3.3. Primjena autorskoga prava: 1.3.3.1. Povijest pregovora 1.3.3.2. Informacije o autorskom pravu 1.3.3.2.1. Definicija autorskoga prava 1.3.3.2.1.1. Ime nakladnika 1.3.3.2.1.2. Nadnevak objavljivanja 1.3.3.2.1.3. Mjesto objavljivanja 1.3.3.2.1.4. Ograničenje pristupa 1.3.3.2.1.5. Osobe za kontakt ili vlasnici autorskoga prava 1.3.3.2.2. Korisnici (koji smiju pristupiti elektroničkoj građi) 1.3.3.2.3. Dopuštene intervencije na građi 1.3.3.2.3.1. Dopuštene intervencije na osnovi postojećega autorskoga prava 1.3.3.2.3.1.1. Poveznica na autorsku izjavu ili zakon 1.3.3.2.3.2. Dopuštena djela s obzirom na licence 1.3.3.2.3.2.1. Poveznica na tekst licence ili druge ugovore
1.4. INFORMACIJA O VJERODOSTOJNOSTI:	1.4.1. Indikator vjerodostojnosti
2. INFORMACIJA O SADRŽAJU:	
2.1. INFORMACIJE O NAČINU PRIKAZIVANJA SADRŽAJA (RI):	2.1.1. Informacija o strukturi: 2.1.1.1. Kratak opis oblika 2.1.1.2. Objekti, koji omogućavaju čitkost dokumenta (<i>transformer objects</i>) 2.1.1.2.1. Platforma 2.1.1.2.2. Parametri 2.1.1.2.3. Strojevi za prikazivanje/analizu (softver) 2.1.1.2.4. Izlazni format 2.1.1.2.5. Unosni format 2.1.1.3. Objekti za prikazivanje/analizu/konverziju 2.1.2. Semantičke informacije: 2.1.2.1. Objekti za prikazivanje/analizu
2.2. PODATKOVNI OBJEKT	Elektronička građa

Iscrpniju analizu o izvodivosti implementacije izbora metapodataka za dugoročnu pohranu iz informacijskog modela OAIS pri arhiviranju elektroničke mrežne građe obavili su partneri **projekta NEDLIB** u okviru razvoja funkcionalnoga modela depozitarnoga sustava elektroničkih publikacija (*Deposit System for Electronic Publications* - DSEP).³³⁵ Jedan od zadataka na projektu bio je određivanje minimalnoga broja metapodataka koji je potreban za dugoročnu pohranu elektroničke građe. Ishodište za odabir metapodataka bile su informacije o načinu prikazivanja sadržaja (*Representation Information*) te informacije koje su važne za dugoročnu pohranu (*Preservation Description Information*).

Dugoročna pohrana znači omogućivanje dugoročnoga pristupa binarnim informacijama, koje su smještene na nekom materijalnom nositelju, metapodatci o načinu prikazivanja sadržaja (RI) tijesno su povezani s informacijama o tehničkim zahtjevima za prikazivanje sadržaja elektroničkih publikacija i omogućavaju njihovo upravljanje. Komponente koje se moraju uzeti u obzir u informacijama o načinu prikazivanja sadržaja (RI) vidljive su iz informacijskoga modela, koji je sastavljen od pet slojeva ili razina (vidi **sliku br. 20**). Svaka razina odgovara jednom određenom aspektu elektroničke građe.

³³⁵ Steenbakkens, J. The Nedlib Guidelines : Setting up a Deposit System for Electronic Publications. Den Haag : Koninklijke Bibliotheek, 2000. (NEDLIB Report series ; 5). URL: <http://www.kb.nl/coop/nedlib/results/NEDLIBguidelines.pdf>. (15.12.2006.)

RAZINA APLIKACIJE → APLIKACIJA (MSWord)	
RAZINA ELEKTRONIČKE GRAĐE → FORMAT (*.doc, *.rtf, Microsoft-Word ...)	
RAZINA STRUKTURE → PODATCI KOJI SU POTREBNI ZA INTERPRETERE I KOMPJLERE	
BINARNA RAZINA → SUSTAV DATOTEKA → OS (Windows NT, XP ... UNIX)	
FIZIČKA RAZINA → FORMAT SKLADIŠTENJA → Automatski generira sustav	HW

Slika br. 20: Informacije o načinu prikazivanja elektroničke građe (representation information)

Materijalna razina - to je standardizirani materijalni nositelj zapisa kojem možemo pristupiti pomoću informacijske tehnologije (hardvera). Metapodatci koji se odnose na taj dio čuvaju se u okviru informacija o izvoru. Metapodatci o hardveru imaju smisla samo u slučaju računalnih programa, kojima trebaju dodatne upute, ili u slučaju multimedijских aplikacija, koje se koriste dodatnim napravama. U svom izvješčaju Lupovici i Masanès (2000, str. 11) navode da će informacije o operacijskim sustavima za 20-50 godina biti neupotrebљive i da je već u okviru katalogizacijskih pravila besmisleno navoditi minimalne zahtjeve za djelovanje određene aplikacije jer znamo da se tehnologija neprekidno razvija;

Binarna razina neovisna je o nositelju i odnosi se na logički binarni zapis i određenu organizaciju toga zapisa u blokovima ili datotekama. Na toj razini djeluju operacijski sustavi koji "poznaju" datotečni sustav. Metapodatci sadrže informacije o tipovima operacijskoga sustava;

Razina strukture - bitovi (*bytes*) su skupљeni u primitivne strukture podataka kojima upravljaju programi. Prevodioci (*compilers* ili kompajleri) i prikazivači (*interpreters*)

programskih jezika omogućavaju funkcionalnost na toj razini. Metapodatci o njima potrebni su ako želimo sačuvati programske aplikacije koje se više ne razvijaju.

Razina elektroničke građe (objekta) - podatci su strukturirani u objekte, koji imaju značenje za aplikaciju i za korisnika. Metapodatci nam daju informaciju o formatu tog objekta.

Razina aplikacije - programske aplikacije (softver) upravljaju objektima iz prethodne razine i predstavljaju ih korisnicima. Metapodatci nam daju informaciju o programskoj aplikaciji i njezinim varijacijama za pristup određenom formatu objekta.

Na projektu NEDLIB ustanovili su da postoje dva tipa elektroničkih publikacija i dokumenata, za koje su nam potrebne dvije vrste metapodataka:

- elektroničke publikacije, koje djeluju na aplikacijama koje su ovisne o specifičnim sustavima; pritom su obavezni metapodatci o aplikacijskoj i binarnoj razini;
- elektroničke publikacije, koje su kreirane u formatima koji su neovisni od specifičnih sustava (XML, JPEG..); u tom slučaju metapodatci o razini objekta nužno su potrebni.

Osim informacija o načinu prikazivanja sadržaja (RI), uzimaju se u obzir i informacije koje su važne za dugoročnu pohranu (PDI). Budući da je projekt NEDLIB prije svega usredotočen na tehnološki aspekt dugoročne pohrane, te informacije podijeljene su na referentne, informacije o vjerodostojnosti (*fixity information*) i informacije o povijesti promjena, koje obuhvaćaju informacije o podrijetlu, kao i informacije o kontekstu.

Na osnovi strukture metapodataka, stručnjaci na projektu NEDLIB definirali su minimalni izbor, koji je potreban za dugoročnu pohranu elektroničke građe. Usredotočili su se na informacijsko-tehnološku stranu i uvjete te zahtjeve sustava za pristup elektroničkim sadržajima u budućnosti. (vidi **tabelu br. 5**).

Tabela br. 5: Izbor metapodataka za dugoročnu pohranu u projektu NEDLIB (Izvor: Lupovici in Masanès, 2000)

1. METAPODATCI O NAČINU PRIKAZIVANJA (<i>Representation Information</i>)	
1.1. SPECIFIČNI ZAHTEVI HARDVERA (HW) - opis nestandardnih konfiguracija platformi ili zahtjeva HW	1.1.1. specifični zahtjevi mikroprocesora 1.1.2. specifični multimedijски zahtjevi 1.1.3. specifični periferni zahtjevi
1.2. OPERACIJSKI SUSTAV	1.2.1. ime 1.2.2. verzija
1.3. PREVODIOČ (<i>interpreter</i>) ILI PRIKAZIVAČ (<i>kompajler</i>)	1.3.1. ime 1.3.2. verzija 1.3.3. upute
1.4. FORMAT OBJEKTA	1.4.1. ime 1.4.2. verzija
1.5. APLIKACIJA	1.5.1. ime 1.5.2. verzija
2. METAPODATCI KOJI SU VAŽNI ZA DUGOROČNU POHRANU SADRŽAJA (<i>Preservation Description Information</i>)	
2.1. REFERENTNE INFORMACIJE	2.1.1. kreator 2.1.2. naslov 2.1.3. nadnevak kreiranja 2.1.4. nakladnik 2.1.5. identifikator 2.1.5.1. vrijednost 2.1.5.2. način oblikovanja 2.1.5.3. odgovorna agencija 2.1.6. URL 2.1.6.1. vrijednost 2.1.6.2. nadnevak upisa
2.2. INFORMACIJA O VJERODOSTOJNOSTI	2.2.1. kontrolni zbroj (<i>checksum</i>) 2.2.1.1. vrijednost 2.2.1.2. algoritam 2.2.2. digitalni potpis
2.3. POVIJEST PROMJENA	2.3.1. važnije promijene izvedene u digitalnom objektu 2.3.1.1. nadnevak 2.3.1.2. stara vrijednost 2.3.1.3. nova vrijednost 2.3.1.4. alat kojim su izvedene promjene: 2.3.1.4.1. ime 2.3.1.4.2. verzija 2.3.1.5. opis postupaka za povratak stare vrijednosti 2.3.2. metapodatci o drugim promjenama 2.3.2.1. stara vrijednost 2.3.2.2. nova vrijednost

Cilj radne skupine OCLC (*Online Computer Library Center, Inc.*) i RLG (*Research Library Group*) bio je određivanje univerzalnoga izbora metapodataka za dugoročnu pohranu. Za osnovu su uzeli dotadašnje definicije metapodataka za dugoročnu pohranu, koje su nastale u okviru ili pod utjecajem referentnoga modela OAIS, i oblikovali sintezu na osnovi četiriju rješenja: CEDARS-a, Nacionalne knjižnice Australije, NEDLIB-a i OCLC-a. Sintezu su dogradili tako što su upotrijebili entitete iz informacijskoga modela OAIS te ih detaljnije specificirali i u hijerarhijskoj strukturi definirali različite vrste informacija na osnovi kojih se stvaraju

metapodatkovni elementi. Rezultati njihovih istraživanja (ožujak 2000. - lipanj 2002.) prikazani su u iscrpnom izvještaju.³³⁶ (vidi **tabelu br. 6**)

Za razliku od europske ili američke prakse, Australci i Novozelanci pragmatično su pristupili određivanju izbora metapodataka za dugoročnu pohranu.

U **australskoj Nacionalnoj knjižnici** u skladu sa selektivnim pristupom odlučuje se o svakoj građi posebno. Određen broj stručnjaka za svaku jedinicu građe temeljito prosuđuje je li potrebno sačuvati je ili ne. Zato je vrlo važno znati tko je odgovoran za donošenje pojedinih odluka i zašto je neka odluka donesena. Zanimljivo je da Australci razlikuju koncept dijela (*work*), koji je konstanta i se odnosi na originalne izvode elektroničke građe, od njegove manifestacije, koja ovisi o strategiji pohrane ili načinu pristupa i prikazivanja s obzirom na postojeću tehnologiju. Pritom se donose odluke je li potrebno sačuvati samo originalno djelo ili i njegovu manifestaciju.

³³⁶ *Preservation Metadata and the OAIS Information Model. A metadata framework to support the preservation of digital objects* / OCLC/RLG Working Group on Preservation Metadata, 2002. URL: http://www.oclc.org/research/projects/pmwg/pm_framework.pdf. (5. 5. 2004.)

Tabela br. 6: Izbor metapodataka OCLC i RLG

1. INFORMACIJE O SADRŽAJU	
1.1. SADRŽAJNI PODATKOVNI OBJEKT	
1.2. INFORMACIJA O NAČINU PRIKAZIVANJA (RI)	
1.2.1. Opis sadržajnog podatkovnog objekta	<ul style="list-style-type: none"> • KRATAK OPIS SADRŽAJA (npr. datotečni sustav, kako su datoteke povezane i kako ih možemo otvoriti) • TIP STRUKTURE (slika, tekst, baza podataka, zvučna snimka ...) • TEHNIČKA INFRASTRUKTURA SLOŽENIH OBJEKATA (nabranje različitih formata datoteka koje čine građu) • OPIS DATOTEKA (tehničke specifikacije svake datoteke) • ZAHTJEVI I POSTUPCI PRI INSTALACIJI • VELIČINA OBJEKTA (u bitovima) • ELEMENTI KOJI SPRJEČAVAJU PRISTUP (enkriptika, lozinke i stalo) • ELEMENTI KOJI OLAKŠAVAJU PRISTUP (sustavi ili metode koji nam omogućavaju pristup podatkovnom objektu, koji treba s vremenom održavati) • BITNE KARAKTERISTIKE KOD PRIKAZIVANJA PODATAKA (boja, oblik, zvuk, struktura ...); • FUNKCIONALNOST (način djelovanja ili prikazivanja) • OPIS PRIKAZANOGA SADRŽAJA (kako se sadržaj mora pregledati i kako ga moraju razumjeti korisnici - sadrži i objašnjenje nejasnih podataka, definicije i opis struktura podataka) • NEDOSTATCI I POGREŠKE - gubitak funkcionalnosti ili promjene u prikazivanju • DOKUMENTACIJA - sva dokumentacija, koja je potrebna za prikazivanje i razumijevanje sadržaja (priručnici, rječnici ...)
1.2.2. Opis okruženja djelovanja	<p>OKRUŽENJE SOFTVERA</p> <ul style="list-style-type: none"> • APLIKACIJE ZA PRIKAZIVANJE <ul style="list-style-type: none"> ○ PROCES TRANSFORMACIJE - softver, potreban za transformaciju bitnog toka u razumljiv sadržaj <ul style="list-style-type: none"> ▪ Alati za transformaciju <ul style="list-style-type: none"> • parametri • format unosa • izlazni format • lokacija alata za transformaciju u sustavu • dokumentacija ▪ Aplikacija za prikazivanje/pristup <ul style="list-style-type: none"> • unosni format • izlazni format • lokacija aplikacije u sustavu • dokumentacija ○ OPERACIJSKI SUSTAV <ul style="list-style-type: none"> ▪ ime ▪ verzija ▪ lokacija u sustavu ▪ dokumentacija <p>OKRUŽENJE HARDVERA</p> <ul style="list-style-type: none"> • lokacija • računalni resursi <ul style="list-style-type: none"> ○ zahtjevi mikroprocesora ○ zahtjevi memorijske jedinice ○ dokumentacija • skladištenje <ul style="list-style-type: none"> ○ informacija o skladištenju ○ dokumentacija • periferne jedinice <ul style="list-style-type: none"> ○ zahtjevi perifernih jedinica ○ dokumentacija

2. INFORMACIJE VAŽNE ZA DUGOROČNU POHRANU GRAĐE (PDI)	
2.1. REFERENTNE INFORMACIJE	2.1.1. IDENTIFIKACIJA ARHIVSKOGA SUSTAVA (LOKALNI IDENTIFIKATOR) <ul style="list-style-type: none"> vrijednost metoda stvaranja odgovorna institucija 2.1.2. OPĆA IDENTIFIKACIJA (bibliografski identifikatori, koje dodjeljuju službene agencije) <ul style="list-style-type: none"> vrijednost metoda stvaranja odgovorne agencije 2.1.3. OPIS ELEKTRONIČKE GRAĐE <ul style="list-style-type: none"> Postojeći metapodatci <ul style="list-style-type: none"> primjer zapisa
2.2. INFORMACIJE O KONTEKSTU	2.2.1. RAZLOZI ZA STVARANJE ELEKTRONIČKE GRAĐE 2.2.2. POVEZANOSTI <ul style="list-style-type: none"> manifestacija <ul style="list-style-type: none"> vrsta povezanosti identifikacija Intelektualni sadržaj <ul style="list-style-type: none"> vrsta povezanosti identifikacija
2.3. INFORMACIJE O PODRIJETLU	2.3.1. KAKO JE GRAĐA BILA STVORENA (skenirana, ili kreirana elektronički ili kako drugačije) 2.3.2. POVIJET GRAĐE OD NASTANKA DO UNOSA U ARHIV 2.3.3. OPIS PROCESA UNOSA GRAĐE U ARHIV 2.3.4. OPIS PROCESA PO UNOSU GRAĐE U ARHIV (održavanje, promjene u sadržaju i druge posljedice dugoročne pohrane) 2.3.5. AUTORSKA PRAVA 2.3.6. AKTIVNOSTI KOJE SU UTJECALE NA GRAĐU <ul style="list-style-type: none"> naziv aktivnosti opis postupka nadnevak odgovorna agencija rezultati postupka opomene
2.4. INFORMACIJE O VJERODOSTOJNOSTI	2.4.1. DOKAZ VJERODOSTOJNOSTI <ul style="list-style-type: none"> metoda ili sustav postupak nadnevak rezultat

U australskoj Nacionalnoj knjižnici uzimaju u obzir razinu agregacije ili granularnosti, koja je bitna za određivanje potrebnih metapodataka za njihovu dugoročnu pohranu građe. To znači da je važno je li elektronička građa zbirka, objekt ili datoteka. Razlikuju se tri razine granularnosti objekta: razina datoteke, razina logičkoga objekta (skup datoteka koje zajedno prikazuju jednu cjelinu) i razina zbirke (skup samostalnih i nepovezanih datoteka). Za pojedinačne vrste datoteka koje imaju zajednička svojstva predviđeni su dodatni metapodatci. To su slike, zvučne snimke, video datoteke, baze i sustavne ili programske datoteke. Povijest promjena može se bilježiti u jednom zapisu ili se za svaku promjenu može otvoriti nov zapis.

Australci preporučuju vrste metapodataka za dugoročnu pohranu elektroničke građe, koji se mogu odnositi na sve ili na neke razine objekata.³³⁷ (vidi **tabelu br. 7**)

Tabela br. 7: Metapodatci za dugoročnu pohranu elektroničke građe u australskoj Nacionalnoj knjižnici

METAPODATCI	OPIS
1. TRAJNI IDENTIFIKATORI;	
2. NADNEVAK KREIRANJA;	
3. TIP STRUKTURE	(tip datoteke ili objekta, koji može biti slika, zvučna snimka, tekst, multimedijaska datoteka ...)
4. TEHNIČKA INFRASTRUKTURA SLOŽENOGA OBJEKTA	sadrži strukturu, odnosno formate datoteka u njegovu sastavu
5. OPIS DATOTEKE	za svaki tip datoteke (slike, zvučne snimke, video datoteke, tekstovi, baze i sustavne ili programske datoteke) predviđeni su određeni atributi, koje opisujemo pomoću metapodataka
6. POZNATI ZAHTJEVI SUSTAVA	operacijski sustav ili okruženje u kojem možemo pristupiti građi u kojoj djeluje
7. ZAHTJEVI ZA INSTALACIJU	bilo koji specijalizirani postupak koji je potreban za instalaciju
8. INFORMACIJA O SKLADIŠTENJU	veličina datoteke ili objekta i lokacija u skladištu
9. ELEMENTI ZA OGRANIČENJE PRISTUPA	sve što ograničava pristup (enkriptika ili digitalni potpis) i koja može utjecati na strategiju za dugoročnu pohranu
10. ALATI ZA TRAŽENJE I PRETRAŽIVANJE TE ELEMENTI ZA OLAKŠAVANJE PRISTUPA (lozinke)	S vremenom treba i te alate održavati da bi se omogućio pristup građi
11. DOPUŠTENJA ZA AKTIVNOSTI KOJA SU POTREBNA ZA DUGOROČNU POHRANU	dopuštenja vlasnika autorskih prava za reprodukciju, migraciju ili konverziju građe u svrhu dugoročne pohrane
12. PROVJERAVANJE VJERODOSTOJNOSTI ILI VALJANOSTI GRAĐE (Validation)	sustav koji nam omogućava provjeravanje je li originalni primjerak promijenjen ili ne; na razini datoteke sve veću važnost dobiva kontrolni zbroj (checksum)
13. POVEZANOSTI	pokazuju je li određena manifestacija građe povezana s ostalima te je li određena građa dio zbirke ili ne
14. GUBITCI I POGREŠKE	svaki gubitak u funkcionalnosti ili izgledu
15. ODLUČIVANJE O ARHIVIRANJU DJELA	odlučivanje o arhiviranju, nadnevak donošenja odluke i nadnevak predviđene revizije odluke
16. RAZLOZI ZA ODLUKU O ARHIVIRANJU GRAĐE	
17. INSTITUCIJA, KOJA JE ODGOVORNA ZA ODLUČIVANJE O ARHIVIRANJU DJELA	
18. ODLUČIVANJE O ARHIVIRANJU MANIFESTACIJE	odlučivanje o arhiviranju, odnosno dokad treba ostati u zalih, te nadnevak odluke i nadnevak predviđene revizije odluke
19. RAZLOZI ZA ODLUKU O ARHIVIRANJU MANIFESTACIJE	
20. INSTITUCIJA KOJA JE ODGOVORNA ZA ODLUKU O ARHIVIRANJU MANIFESTACIJE	
21. PREDVIĐENA UPOTREBA ODREĐENE MANIFESTACIJE	
22. INSTITUCIJA KOJA JE ODGOVORNA ZA DUGOROČNU POHRANU	
23. PROCES	svi relevantni podaci o bilo kojem procesu koji je izveden u vezi s digitalnim objektom ili datotekom
24. KREATOR ZAPISA	
25. OSTALO	bilo koja druga relevantna informacija o elektroničkoj građi

Kao što vidimo u **tabeli br. 7**, Australci su se usredotočili na informacije izvan sustava koje nam trebaju za upravljanje dugoročnom pohranom. Metapodatci su jednako upotrebljivi i za elektroničku građu koja je nastala u elektroničkom obliku, kao i za digitalne kopije originala na papiru. Ne bave se procesima koji su potrebni za dugoročnu pohranu (migracija ili emulacija).

³³⁷ *Preservation metadata for digital collections* : exposure draft. Canberra: National Library of Australia, 1999. URL: <http://www.nla.gov.au/preserve/pmeta.html>. (5. 5. 2004)

Popis metapodataka neovisan je o tehnologiji te stoga ne prati sasvim model OAIS. Njihova strategija dugoročne pohrane temelji se na analizi radne skupine OCLC/RLG iz 2001. godine, iako je australski pristup mnogo više pragmatičan i neposredno povezan s praksom.

Izbor metapodataka za dugoročnu pohranu elektroničke građe koji su pripremili u **novozelandskoj Nacionalnoj knjižnici** vrlo je bitan za daljnji razvoj ovoga područja.³³⁸ U njihovim smjernicama i standardnim okvirima, za razliku od ostalih nacionalnih knjižnica, vrlo se jasno razlikuju originalni primjerak građe (*preservation masters*) i ostale verzije koje su priređene za pristup, odnosno za diseminaciju ili upravljanje. Predmet njihove dugoročne pohrane su osuvremenjene verzije originalnoga primjerka građe te prvobitni nepromijenjeni originali u izvornom formatu (naprimjer originalna izdanja na CD-ROM-u), koji već mogu biti nečitki zbog razvoja informacijske tehnologije. Osim toga, pri odlučivanju građa koja je nastala u elektroničkom obliku ima prednost pred građom koja je rezultat konverzije iz analognih ili materijalnih medija u digitalne (naprimjer skenirane slike ili tekstovi).

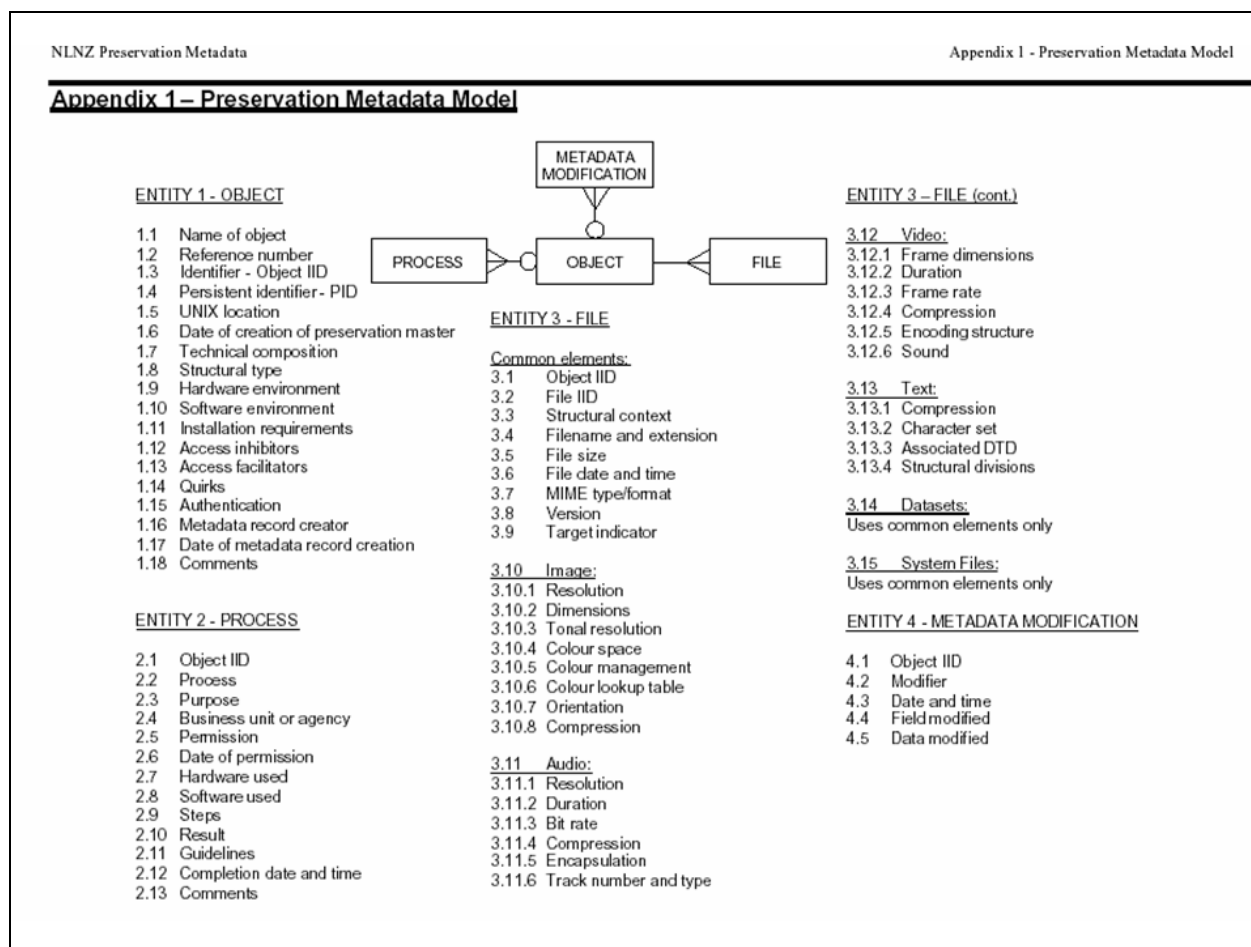
Po uzoru na australsku Nacionalnu knjižnicu, Novozelanci također obraćaju pažnju stupnju granularnosti digitalnih objekata i dijele ih na jednostavne, koji se sastoje od jedne datoteke, složene datoteke, koji se sastoje od skupine nesamostalnih datoteka koje zajedno čine jedan logički objekt (naprimjer mrežna stranica ili baza podataka), i **skupine objekata** koje se sastoje od skupine samostalnih datoteka koje nisu ovisne jedna o drugoj (naprimjer disketa sa dvjesto pisama poznatoga pisca u WordPerfectu i Wordu).

Njihov model obuhvaća četiri entiteta koji su međusobno povezani (**slika br. 21**):

- **Objekt** - odnosi se samo na originalni primjerak logičkoga objekta koji može biti jednostavan, složeni objekt ili skupina objekata; objekt definira osamnaest vrsta metapodataka koji su važni za dugoročnu pohranu; među njima su deskriptivni metapodatci, metapodatci koji opisuju tehničke zahtjeve za pristup građi, odnosno za njezino djelovanje, informacijsko tehnološko okruženje, funkcionalnost građe i način pristupa;

³³⁸ *Metadata standards framework - preservation metadata* (2002). Wellington: National Library of New Zealand, 2002. URL: http://www.natlib.govt.nz/files/4initiatives_metaschema.pdf. (15.12.2006.)

- **Proces** - to je bilo koja aktivnost koja je izvedena nad originalnim primjerkom elektroničke građe i evidentira se pomoću trinaest elemenata ili metapodataka; među njima je i razlog zašto je te vrijeme kada je aktivnost izvedena, onaj tko ju je odobrio, način na koji je izvedena, opis upotrijebljene opreme za taj proces i konačno rezultati tih promjena;
- **Datoteka** - sadrži tehničke informacije o najnižim razinama digitalnih objekata; devet elemenata zajedničko je svim tipovima datoteka (ime datoteke, ekstenzija, nadnevak kreiranja, format ...); specifična svojstva odnose se na slike, zvučne snimke, video datoteke i tekstove.
- **Promjene metapodataka** - evidentira sve promjene koje su izvedene na zapisu sa metapodacima pomoću pet elemenata.



Slika br. 21 : Izbor metapodataka za dugoročnu pohranu elektroničke građe u Novom Zelandu (vidi: *Metadata standards framework - preservation metadata* (2002). Wellington: National Library of New Zealand. URL: http://www.natlib.govt.nz/files/4initiatives_metaschema.pdf , str. 26)

Utjecaj novozelandskoga izbora osjetio se u OCLC-u i RLG-u te u njemačkoj Nacionalnoj knjižnici.

OCLC i RLG su 2003. godine osnovali međunarodnu skupinu PREMIS (*Preservation Metadata: Implementation Strategies*) čiji je cilj, između ostalog, bio određivanje skupine osnovnih elemenata metapodataka sa širokom upotrebom pri dugoročnoj pohrani elektroničke građe, izrađivanje "rječnika podataka" (*Data Dictionary*) s izborom metapodataka, ocjenjivanje različitih alternativnih strategija za kodiranje, smještanje i upravljanje metapodacima za dugoročnu pohranu u elektroničkom sustavu za dugoročnu pohranu i ostalo.

U svibnju 2005. godine u svom su izvještaju objavili popis svih mogućih metapodataka za dugoročnu pohranu, koji bi se mogli dodati definiciji elektroničke građe.³³⁹ Za razliku od deskriptivnih metapodataka, metapodatci za dugoročnu pohranu čuvaju se zajedno s elektroničkom građom, po mogućnosti u XML-u.

PREMIS je nastavio rad OCLC-a i RLG-a iz 2002. godine, tražeći način da teorijski rad iz 2002. godine provede u praksi.³⁴⁰ Međutim, dok su OCLC i RLG prije detaljno pratili model OAIS, ovdje se prilično udaljavaju, a koriste i različitu terminologiju te metapodatke specificiraju mnogo detaljnije od OAIS-a. U PREMIS-ovu rječniku elementi metapodataka nazivaju se semantičkim jedinicama jer su to informacije, a ne zapisi. Svaka semantička jedinica vezana je za jednu od pet sljedećih entiteta:

- **objekt** ili digitalni objekt – to je odvojena jedinica informacije u digitalnom obliku; sastoji se od tri podvrste entiteta: datoteke, bitnoga toka (*bitstream*) i načina prikazivanja;³⁴¹
- **intelektualni entitet** - to je povezan skup sadržaja koji se smatra jedinicom građe (naprimjer knjiga, mapa, fotografija, baza podataka, mrežno mjesto itd.); jedan intelektualni entitet može

³³⁹ Data Dictionary for Preservation Metadata : Final Report of the PREMIS Working Group. Dublin (OH) : OCLC ; Mountain View (CA): RLG, 2005.

³⁴⁰ *Preservation Metadata and the OAIS Information Model. A metadata framework to support the preservation of digital objects* / OCLC/RLG Working Group on Preservation Metadata, 2002. URL: http://www.oclc.org/research/projects/pmwg/pm_framework.pdf. (5. 5. 2004.)

³⁴¹ PREMIS koristi termin "prikazivanje" da bi izbjegao upotrebu termina "manifestacije", koji se koristi u modelu Functional Requirements for Bibliographic Records (FRBR)

obuhvaćati druge intelektualne entitete (naprimjer mrežno mjesto može sadržavati jednu mrežnu stranicu, mrežna stranica jednu fotografiju itd.);

- **dogadjaj** (*event*) – to je djelovanje koje uključuje bar jedan objekt ili uzročnika (*agent*) koji je poznat u repozitoriju za dugoročnu pohranu;
- **uzročnik** (*agent*) - to je osoba, organizacija ili softverski program povezan s događajima u vezi sa dugoročnom pohranom u životnoj dobi objekta;
- **Prava** ili navođenje prava – to su izjave autorskih dopuštenja u vezi s objektom i/ili uzročnikom.

Entiteti međusobno ulaze u različite odnose. Intelektualna je jedinica izvan dosega jer nju obrađuju deskriptivni metapodatci. Semantičke su jedinice svojstva tih entiteta (naprimjer, veličina je svojstvo entiteta objekt) i mogu biti kontejner za ostale semantičke jedinice.

PREMIS se ne upušta u opisivanje karakteristika entiteta niti metapodataka koji su već dostupni preko kataloga ili nisu povezani s dugoročnom pohranom elektroničke građe.

U njemačkoj **Nacionalnoj knjižnici** otprilike u isto su vrijeme kao i PREMIS, u travnju 2005. godine, izdali izbor metapodataka *Long-term preservation metadata for electronic resources* (LMER).³⁴² Iako su nastojali da rječnik metapodataka PREMIS bude praktičan i upotrebljiv u praksi dugoročne pohrane, još uvijek je prilično kompliciran i zahtjevan. Nijemci su pratili aktivnosti u PREMIS-u, a također su proučili i praksu Nacionalne knjižnice u Novom Zelandu. Odlučili su primijeniti novozelandsku praksu. Tako LMER dijeli metapodatke za dugoročnu pohranu na četiri dijela koji odgovaraju novozelandskim entitetama: objekt, proces, datoteka i promjene metapodataka. Metapodatke unose u shemu XML, naprimjer METS.

Svi opisani izbori metapodataka mogu se koristiti za svu elektroničku građu. Pri automatskom prikupljanju mrežne građe vrlo je teško ručno dodavati metapodatke i analizirati njihov sadržaj. Može se uključiti robot za prikupljanje tako da većinu tehničkih i administrativnih metapodataka automatski zapisuje u samu građu. Osim toga, većina mrežnih mjesta sastavljena je od različitih datoteka.

³⁴² *Long-term preservation metadata for electronic resources* (LMER). / uredio Steinke, T. Frankfurt am Main : Die Deutsche Bibliothek, 2005. URL: <http://www.ddb.de/eng/standards/lmer/lmer.htm>. (15.2.2007.)

U okviru **konzorcija IIPC** pokušali su odrediti koji će metapodatci biti važni za pristup elektroničkoj građi u budućnosti. Pripremili su sljedeći izbor metapodataka:³⁴³

1) Podatci o dokumentima na mreži:

- lokacija - originalna lokacija i lokacija u arhivu
- veličina
- format (ime, verzija, alat za verifikiranje)
- kontrolni zbroj (*Checksum*)
- predviđanje učestalosti promjena
- vremenska dimenzija
- kontekst u kojem je građa nastala

2) Podatci o činiteljima pri preuzimanju građe na mreži:

- podatci o alatu za prikupljanje mreže (tip, nastavljajanje, IP)
- podatci o poslužitelju (IP, informacija o DNS)

3) Podatci o procesu prikupljanja:

- vrijeme preuzimanja
- zahtjevi klijenta
- odgovor poslužitelja
- status *robots.txt*
- mandat za pristup (SSL)

4) Podatci o procesu odabira:

- početni popis URL-a
- put (*path*) do dokumenta
- preskok iz ulazne točke (ulaznog popisa URL-a)
- dokumentacija politike odabira

³⁴³ Masanès, J. IIPC Web Archiving Toolset. Netpreserve.org // IAW 2005. URL: <http://www.iaw.net/05/masan2.pdf> (15.12.2006.)

- karakterizacija/evaluacija (vrijednost, alat)

5) Povijest promjena

- metapodatci
- nova vrijednost
- alat/aplikacija

10.1. Trajni identifikatori

Identifikatori su metapodatci koji jednoznačno identificiraju jedinicu građe. Kod građe na materijalnim nositeljima poznati su bibliografski identifikatori ISBN (*International Standard Book Number*), ISSN (*International Standard Serial Number*), ISMN (*International Standard Music Number*), SICI (*Serial Item and Contribution Identifier*) i ostali, koji identificiraju odgovarajuće vrste publikacija. Međutim, u digitalnom okruženju oni se vrlo malo upotrebljavaju. Standard za identifikaciju jedinice mrežne građe je URI (*Universal Resource Identifier* = Univerzalni identifikator građe³⁴⁴) koji je sastavljen od imena sheme (primjerice HTTP), imena domene (www.cisco.com) i adrese na samom poslužitelju (eu/partners/index.html).³⁴⁵ Najčešći oblik URI-a je URL (*Uniform Resource Locator* = jedinstvena adresa građe)³⁴⁶ i označava poslužitelja na kojem je građa smještena.³⁴⁷ Osim HTTP protokola (*hypertext transfer protocol services*) sheme URI postoje i za druge protokole, kao što su FTP (*File Transfer Protocol services*), Gopher shema i protokol i mailto shema za elektroničke naslove i poštu.

³⁴⁴ Berners-Lee, T. ; Masinter, L. ; McCahill, M. RFC 1630 - Universal Resource Identifiers in WWW (URI), 1994. URL: <http://tools.ietf.org/html/rfc1630>. (3.11.2006)

³⁴⁵ Scott-Wilson, E. Identifiers and interoperability // Information architecture: designing information environments for purpose / uredili Glichrist, A. ; Mahon, B. London : Facet, 2004. (Managing information for the knowledge economy series / Series Editor: Angela Abell). Deseto poglavlje, str. 161-173.

³⁴⁶ Berners-Lee, T., Masinter, L. and McCahill, M.: RFC 1738 - Uniform Resource Locators (URL), 1994. URL: <http://rfc.dotsrc.org/rfc/rfc1738.html>. (3.11.2006)

³⁴⁷ Identifikatore je prvi upotrijebio osnivač svjetske mreže, Tim Berners-Lee, koji je izradio prijedlog za njegovu strukturu i objavio ga na mreži kao RFC (*Request for comments*= zahtjev za komentare). URI obuhvaća različite oblike, među njima su najpoznatiji URL i URN.

Problem nastaje kad građa promijeni smještaj i zbog "prekinute" poveznice ona više nije dostupna. Kao rješenje uvedeni su trajni identifikatori koji se za razliku od URL-a dodjeljuju građi, a ne smještaju je.³⁴⁸ Trajni identifikatori vezani su uz baze podataka sa sustavima razlučivanja (resolveri) koji se trenutačno osuvremenjuju, a u kojima se identifikatori vežu za URL te tako preusmjeravaju korisnika do aktivnih adresa mrežne građe.

Konzorcij OCLC uveo je sustav PURL (*Persistent Uniform Resource Locator* - URL, <http://purl.oclc.org/>) koji omogućava trajno preusmjeravanje na jedinicu mrežne građe. Funkcionalno, PURL je zapravo običan URL, sa tom razlikom da između PURL-a i pravog URL-a postoji standardno HTTP automatsko 'preusmjeravanje' na aktivnu adresu na mreži.³⁴⁹ Poslije su prihvaćene ostale vrste identifikatora, koji su omogućili da se jedinica građe pronalazi preko različitih sustava razlučivanja.³⁵⁰ Među njima je i URN (*Uniform Resource Name* = jedinstveni naziv građe), podvrsta URI-a, o kojem će biti govora poslije.

Potreba za koordinacijom procesa dodjeljivanja identifikatora i njihovo centralizirano vođenje te osiguravanje sigurnoga mrežnoga pristupa elektroničkim dokumentima bio je glavni uzrok za osnivanje agencija, koje su se financirale članarinama svojih članica. Te agencije brinu za sljedljivost digitalnih objekata kojima su dodijelile identifikatore i na neki način jamče za njihovu vjerodostojnost, cjelovitost i originalnost.

³⁴⁸ O podjelama i vrstama identifikatora vidi: Reynolds, R. ISSN, identifikatori i metapodatci u digitalnom svijetu = ISSN, identifiers and metadata in the digital world. // Vjesnik bibliotekara Hrvatske 47, 1-2(2004), str. 16-24. Regina Reynolds dijeli identifikatore na "aktivne" i "trajne", te na "inteligentne" i "neinteligentne". Aktivni su naprimjer URL-i, koji označavaju samo smještaj građe, dok trajni sadrže rješenje problema prekinutih poveznica, jer se dodeljuju jedinici građe i omogućuju njezino pronalaženje bez obzira na poslužitelja na kojem je smještena. Primjer trajnih identifikatora navodi URN i DOI. Inteligentni identifikatori sadrže u sebi neke podatke o jedinici građe. Takav je ISBN, jer iz njegove strukture možemo ustanoviti u kojoj zemlji ili području je elektronička knjiga objavljena, saznajemo o identifikatoru nakladnika, stvarnog naslova i kontrolnog broja koji štiti od pogrešaka u prepisivanju. Neinteligentni identifikatori ne sadrže nikakve informacije, već ih na osnovi tog broja moramo tražiti drugdje, obično u agencijama, koje vode te registre. Primjer neinteligentnoga identifikatora je ISSN. Po Reynoldsovoj podjeli možemo zaključiti da trajni identifikatori mogu biti inteligentni ili neinteligentni. Inteligentni trajni identifikatori, primjerice URN-i ili ARK, uključuju bibliografske identifikatore ili informacije o jedinici građe, dok su neinteligentni trajni identifikatori često formirani na osnovi metode "*numerus currens*" brojevnoga algoritma neke agencije, koja ih dodjeljuje, ili na temelju kontrolnoga zbroja (*checksuma*) odnosno MD5 koda elektroničke građe.

³⁴⁹ Scott-Wilson, E. Identifiers and interoperability. // Information architecture: designing information environments for purpose / uredili Glichrist, A. ; Mahon, B. London : Facet, 2004. Str. 163.

³⁵⁰ Više o trajnim identifikatorima može se naći u: Hilse, H.W. ; Kothe, J. Implementing persistent identifiers : overview of concepts, guidelines and recommendations. London : Consortium of European Research Libraries (CERL) ; Amsterdam : European Commission on Preservation and Access (ECPA), 2006.

Takvi su sustavi Handle, koji dodjeljuje *Corporation for National Research Initiatives*, i ARK³⁵¹ iz kalifornijske Digitalne knjižnice. Dobivanje tih dvaju identifikatora ovisi o članarini. Sustavima Handle i ARK zajedničko je to da se upisuju kao produžetak URL-a. Znači, prvi dio upućuje na smještaj građe na mreži, i taj se dio može promijeniti, dok drugi dio, koji je slobodno generiran broj, često počinje kodom hdl. ili ark.

Iz sustava Handle nastao je DOI (*Digital Object Identifier* = identifikator digitalnog objekta), koji dodeljuje *International DOI Foundation* (Međunarodna zaklada za DOI). DOI je administrativni okvir za dodjeljivanje trajnih identifikatora i prihvaćaju ga elektronički nakladnici. Koristi se sustavom Handle za dodjeljivanje imena i kao resolver. DOI je brojevnoslovni naziv koji identificira digitalni sadržaj prema intelektualnom vlasniku digitalnoga objekta. Njegova sintaksa čini otvorenu normu (američki Nacionalni standard ANSI/NISO Z39.84-2000). Njegova se vrijednost može usporediti s ISBN-om u tradicionalne građe. Baza DOI sadrži osnovne bibliografske metapodatke o digitalnom objektu. Mnoge europske nacionalne knjižnice prihvatile su ovaj tip identifikatora. Europska konferencija ravnatelja nacionalnih knjižnica (*Conference of European National Librarians* - CENL) postala je članica njihova upravnog odbora³⁵² i u okviru posebne radne skupine o trajnim identifikatorima razmatra njegovu upotrebu u nacionalnim knjižnicama. Njemačka Nacionalna knjižnica istovremeno se koristi sustavima DOI i URN i za to ima odgovarajući sustav razlučivanja (*resolver*).³⁵³

U nastavku je dan primjer opisanih identifikatora:

- **URL:** <http://www.cisco.com/en/US/partners/index.html>
- **URN:** urn:nbn:fi:tkk-004781
- **PURL:** <http://purl.oclc.org/keith/home> --> <http://home.columbus.rr.com/shafer/>
- **Handle:** hdl:4263537/4069 → <http://hdl.handle.net/4263537/4069>
- **DOI:** doi:10.1000/182 → <http://dx.doi.org/10.1000/182>
- **ARK:** <http://foobar.zaf.org/ark:/12025/654xz321/s3/f8.05v.tiff>

³⁵¹ Kunze, J. A. Towards electronic persistence using ARK identifiers, 2002. URL: <http://www.cdlib.org/inside/diglib/ark/arkcdl.pdf>. (3.11.2006)

³⁵² CENL: http://www.nlib.ee/cenl/working_groups.php. (15.12.2006)

³⁵³ Više o tome na mrežnoj strani: <http://nbn-resolving.de/>. (15.12.2006)

U većini nacionalnih knjižnica primjenjuje se sustav URN jer ne zahtijeva članstvo u nekoj agenciji i svaka je knjižnica slobodna da sama formira svoju sintaksu i prati elektroničku građu na nacionalnoj razini. Za razliku od ostalih identifikatora, URN može sadržavati dodatne informacije o jedinici mrežne građe, kao što su lokacijske informacije (signature ili slično) ili bibliografski identifikatori.

Postoje dva načina tvorbe URN-a:³⁵⁴

1. Mogu se umjetno tvoriti na osnovi nekoga tekućega broja i šifre jedinice građe:

URN:nbn:[koda_države]:[akronim_ili_koda_organizacije]:[šifra_tipa_građe/umjetno_stvo
ren_broj_jedinice_građe].

Primjer: URN:nbn:si:NUK:a22839403

2. Mogu se tvoriti na osnovi postojećih podataka (formalno dodijeljenih bibliografskih identifikatora i koda automatski generiranih brojeva, kao što su kontrolni zbroj (*checksum*) ili MD5.

Primjer: URN:NBN:SI:NUK:ISBN:961-6162-83-7<PDF>A0.1 ili

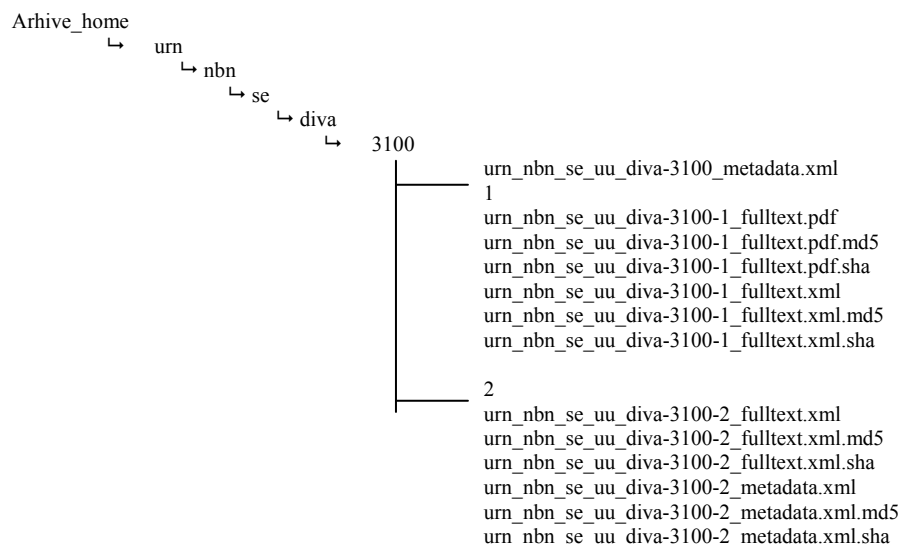
URN:NBN:SI:NUK:p:8F5F7C1108200A26F1E14F02FE7208C8<PDF><20060214>

Umjetno stvoreni URN-i imaju prednost da se svi nastavljaju na jednaku shemu, kraći su i lakše je upravljati njima. Naprimjer, u švedskoj Sveučilišnoj knjižnici u Uppsali izradili su sustav za

³⁵⁴ Više informacija o sintaksi URN-jev u:

- RFC 2141: URN Syntax: <http://www.faqs.org/rfcs/rfc2141.html> <<http://www.ietf.org/rfc/rfc2141.txt>>. (15.12.2006)
- RFC 3044 - Using The ISSN (International Serial Standard Number) as URN (Uniform Resource Names) within an ISSN-URN Namespace: <<http://www.faqs.org/rfcs/rfc3044.html>>. (15.12.2006)
- RFC 3061 - A URN Namespace of Object Identifiers: <<http://www.faqs.org/rfcs/rfc3061.html>> (15.12.2006)
- RFC 3187 - Using International Standard Book Numbers as Uniform Resource Names: <<http://www.faqs.org/rfcs/rfc3187.html>>. (15.12.2006)
- RFC 3188 - Using National Bibliography Numbers as Uniform Resource Names: <<http://www.faqs.org/rfcs/rfc3188.html>>. (15.12.2006)
- RFC 2288 - Using Existing Bibliographic Identifiers as Uniform Resource Names: <<http://www.faqs.org/rfcs/rfc2288.html>>. (15.12.2006)

dodjeljivanje URN-a u njihovu arhivu DiVA. Trude se da održe jednak sustav identificiranja građe u cijelom sustavu. Tako se URN:NBN koristi kao dio konvencije za imenovanje datoteka i direktorija u samom arhivu.³⁵⁵ Primjer je dan na **slici br. 22**.



Slika br. 22: URN:NBN upotrijebljen je kao dio konvencije imenovanja datoteka i direktorija u arhivu DiVA

Nedostatak je taj da se ti trajni identifikatori moraju tvoriti centralizirano pomoću računalne aplikacije (generatora), a potrebno je i vođenje posebnoga registra. Osim toga, iz samog identifikatora nije vidljiv sadržaj elektroničke građe i zato je teža kontrola duplikata.

Prema podacima iz 2004. godine,³⁵⁶ sustav URN:NBN osim u švedskoj koriste i u norveškoj i njemačkoj nacionalnoj knjižnici. U većini ostalih europskih nacionalnih knjižnica razmatraju njegovo uvođenje.

³⁵⁵Müller, E. ... [et al.] Archiving Workflow between a local repository and the national archive. Experiences for the DiVa Project. URL: http://epc.ub.uu.se/files/archiving_ECDL_2003.pdf. (15.12.2006)

³⁵⁶ Summary of the questionnaire "Application of persistent identifiers (PI) at national libraries", Die Deutsche Bibliothek, 26.1.2004. Istraživanje obavljeno u okviru projekta EPIKUR. URN: <http://www.urn-info.org/?link=300>. (15.12.2006)

11. IZGRADNJA METODOLOŠKOGA MODELA ZA POHRANU SLOVENSKE MREŽNE I ELEKTRONIČKE GRAĐE

U prošlim poglavljima prikazali smo različite pristupe pri arhiviranju mreže te prednosti i slabosti tih pristupa. Strogo selektivne metode ograničile bi prikupljanje na ona mrežna mjesta koja su nam danas važna, a ne bi obuhvatile ona mjesta koja bi mogla biti interesantna budućim korisnicima. Naravno, i sam bi odabir pružao dovoljno informacija o kulturi, znanosti i ostalim aspektima društva iz vremena u kome je građa prikupljena. S druge strane, imamo jedinice građe koje tek dobivaju formalno, strukturno i, možemo reći, žanrovsko značenje i ako ih ne skupimo, u budućnosti će se teško rekonstruirati njihov nastanak i razvoj. Iz toga proizlazi da bi oba pristupa zajedno donijela optimalne rezultate.

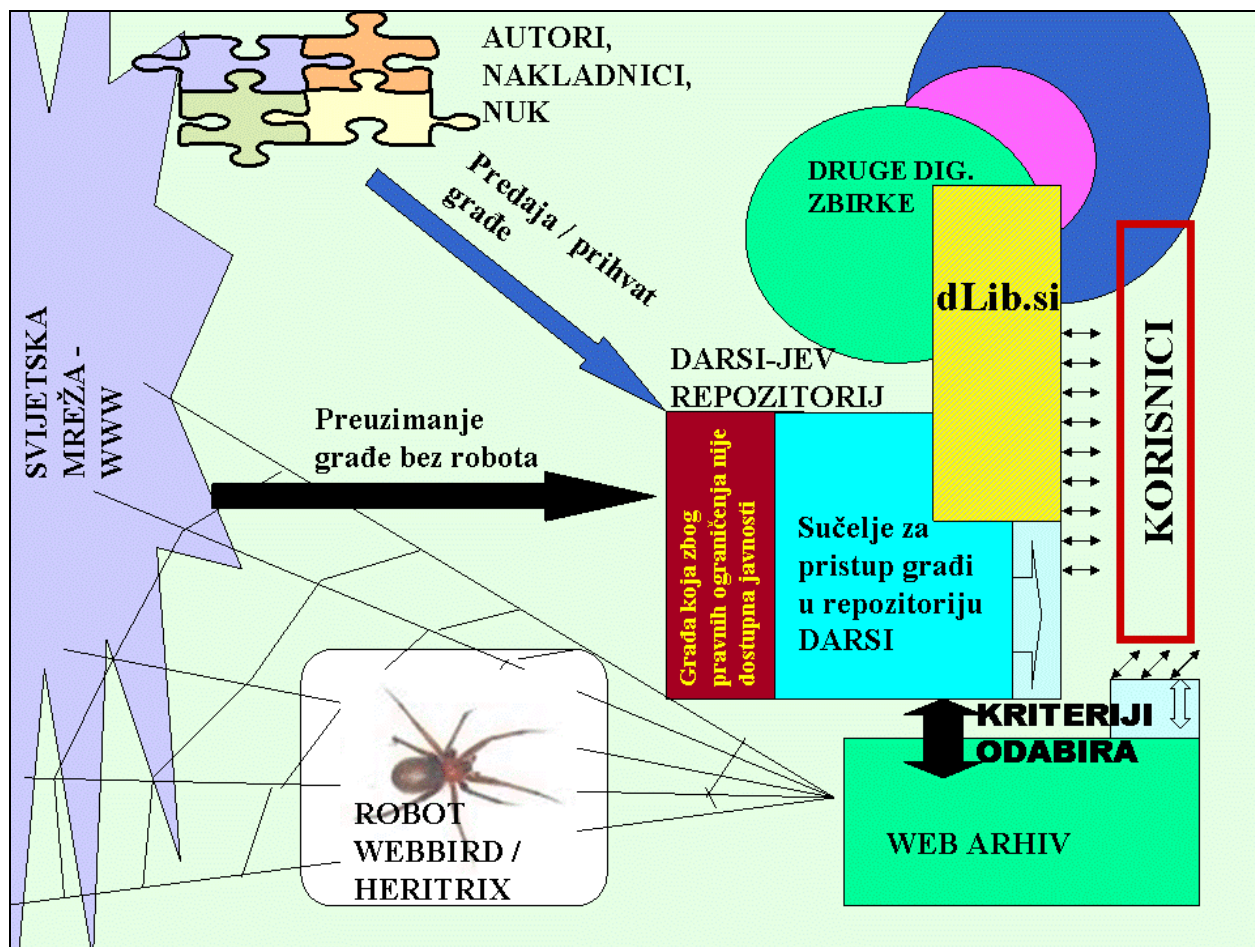
Za NUK je optimalna kombinacija strategija automatskoga i selektivnoga pristupa. Automatskim preuzimanjem uspjela bi se preuzeti sva građa na određenim domenama, ne samo na *.si, već i ostalima na *.org, *.com, *.net itd., koje su na slovenskom jeziku, ili stranice na određenim URL-ima. Dodatno bi se sa svakim nakladnikom koji objavi neki oblik formalne mrežne građe ostvario dogovor o pojedinačnom preuzimanju. Među nakladnicima su akademske institucije koje objavljuju obrazovnu građu na mreži, komercijalni nakladnici koji objavljuju elektroničke serijske publikacije i elektroničke knjige, različite udruge itd. Budući da postoji interes da te publikacije dobijemo bez zaštite i u cjelini te da već na početku prikupljanja dobijemo potrebna dopuštenja za javni pristup, nužna je suradnja s njihovim nakladnicima. Pojedinačno i izravno od nakladnika preuzimala bi se i relevantna dinamična građa koju današnji roboti ne mogu preuzeti.

Po potrebi bi se retrogradno pregledao web arhiv i ako se pokaže potreba za tematsko ili formalno arhiviranje dodatne građe, onda bi se ta građa mogla posebno obraditi i smjestiti u digitalni arhiv. Ako ta građa nije potpuno zahvaćena, pokušalo bi se stupiti u kontakt s njezinim autorima ili nakladnicima.

Za kombiniran pristup potrebna su dva kontejnera, to znači, web arhiv, koji se stvara automatskim preuzimanjem mreže pomoću robota i digitalni repozitorij, u kome su procesi prilagođeni sustavu OAIS. U konzorciju IIPC pokušavaju prenijeti koncept OAIS također u web

arhiv. To je već moguće s formatom WARC, koji se još razvija. Trenutačno je optimalno rješenje preuzimanje mreže s Heritrixom i smještanje mrežnih stranica u XML strukturu, što omogućava vođenje metapodataka o procesima pohrane i konverzije u skladu sa strategijom za dugoročnu pohranu.

Kombinacijom ovih dvaju pristupa uspjeli bismo sačuvati slovensku mrežu. Koncept arhivskoga sustava prikazan je na **slici br. 23**. NUK-ov informacijski portal do digitalne građe je dLib.si. Digitalni repozitorij DARSİ (*Digitalni ARhiv SlovenIje = Digital ARchive of SlovenIa*) sadržavao bi tako građu koja je namijenjena javnosti, a bila bi dostupna preko portala dLib.si, kao i građu koja zbog autorskih prava ne može biti javno dostupna. Web arhiv mogao bi biti dostupan neposredno ili preko portala dLib.si. Prihvat elektroničke građe obavlja se na tri načina: tako da je automatski prikupljaju roboti, tako da je prikupljaju autori i nakladnici te i odabirom iz Web arhiva. U sva tri slučaja postoje kriteriji odabira. Za građu koju primamo od autora i nakladnika kriteriji su formalni i sadržajni, dok je za namještanje robota potrebno postaviti tehničke kriterije zahvaćanja željenih sadržaja.



Slika 23: Koncept arhivskog sustava u NUK-u

Osnovni model procesa temeljio bi se na OAIS modelu (vidi sliku br. 24). Glavni procesi bi bili sljedeći:

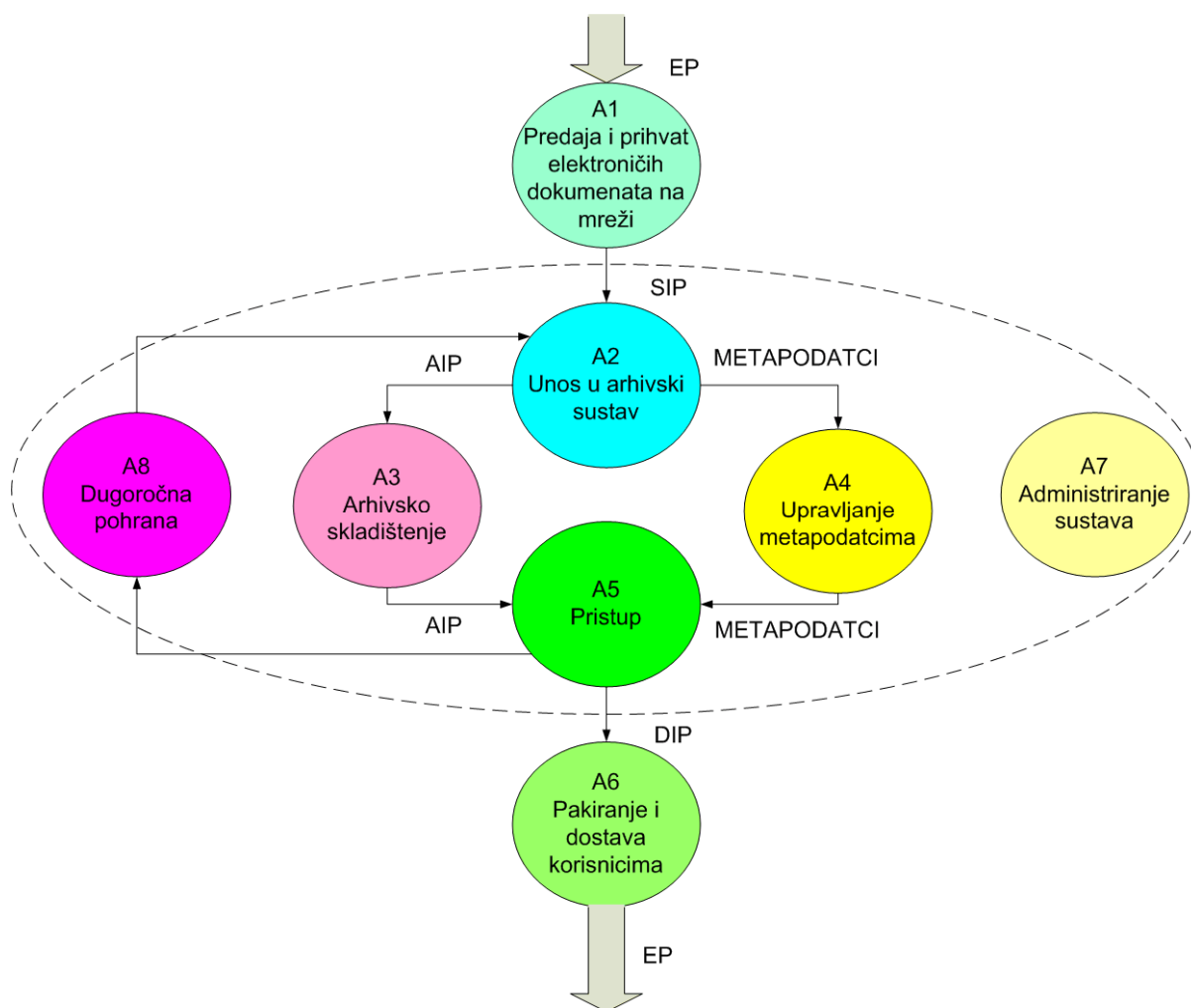
- A1: Predaja i prihvati elektroničkih dokumenata na mreži
- A2: Unos u arhivski sustav
- A3: Arhivsko skladištenje
- A4: Upravljanje metapodacima
- A5: Korisnički pristup
- A6: Pakiranje i dostava građe korisnicima
- A7: Administriranje sustava.
- A8: Dugoročna pohrana.

U nastavku će se detaljnije opisati svaki od spomenutih procesa. Međutim, prije početka bilo kakvog procesa, u vezi s dugoročnom pohranom potrebno je definirati kontekst ili okruženje u kojem se obavljaju ti procesi. Ta okruženja možemo definirati kao organizacijsko okruženje, proceduralne aspekte izvođenja svih procesa i svu dokumentaciju koja podupire procese dugoročne pohrane građe.

Organizacijsko okruženje. Podrazumijeva kontekst i uvjete pod kojima se obavljaju arhivski procesi. Poželjna je izrada organizacijske dugoročne i kratkoročne strategije koja sadrži sljedeće informacije:

- ovlaštenja institucije na području dugoročne pohrane elektroničke građe,
- zakonske osnove za obavljanje tih procesa,
- odgovornosti i suradnja stručnjaka ili odjela u tim procesima,
- mogu biti opće opisane metode i pristupi kod gradnje i održavanja arhivskoga sustava,
- ciljevi koje se želi postići s dugoročnom pohranom,
- sukcesijski plan.

Posebice je preporučljivo izraditi financijski plan za održavanje svih procesa, što znači stručnjaka i infrastrukture.



Slika br. 24: Prikaz procesa u arhivskom sustavu. Po: Werf, T. van der: The Deposit System for Electronic Publications : a process model. - The Hague : NEDLIB Consortium, 2000. URL:

Proceduralni aspekti izvođenja svih procesa - Sve procese (A1-A8) i postupke treba posebice dokumentirati. Posebno bi se morale opisati sljedeće procedure:

- politika i kriteriji odabira građe,
- postupci nabave i preuzimanja građe,
- postupci rješavanja autorskih prava,
- metode obradbe građe,
- navođenje i opis važnih standarda koji podupiru sve aktivnosti,
- metode održavanja i administriranja arhiva i njegove infrastrukture,
- politika i načini omogućavanja pristupa građi korisnicima,
- postupci u slučaju da autori/nakladnici ne žele dopustiti pristup građi,

- postupci u slučaju prirodnih i prouzrokovanih nesreća,
- odgovorni odjeli i osobe u svakom od navedenih postupaka ili procesa.

Svi postupci moraju biti poduprti sa smjernicama, koje detaljnije definiraju sve aktivnosti.

Dokumentacija koja podupire procese dugoročne pohrane građe

Naravno i dokumentacija o procesima dugoročne pohrane može biti napisana u obliku smjernica:

- opis odgovornosti i ovlaštenja odjela ili stručnih osoba za područje dugoročne pohrane,
- opis postupaka za osiguravanje dugoročne pohrane elektroničke građe,
- strategije dugoročne pohrane građe,
- postupci obradbe građe i definicija metapodataka za dugoročnu pohranu,
- definicija i implementacija važećih standarda koji se koriste kao podrška tim strategijama (bibliografski opis, metapodatci, identifikatori, način smještanja u arhiv, standardi formata, postupaka i dugoročne pohrane itd.),
- analiza trajnosti materijalnih nositelja elektroničkih zapisa,
- stvaranje arhivskih kopija i lokacija za njihovu pohranu,
- primjeri i upute za sređivanje autorskih ugovora sa nakladnicima i autorima,
- načini rješavanja oštećene elektroničke građe - povrat podataka,
- načini i postupci stvaranja duploga arhiva te detaljan plan sukcesije,
- koordinacija svih djelatnosti na području dugoročne pohrane i sudjelovanje s drugim depozitarnim ili arhivskim organizacijama, također privatnim, na tom području,
- suradnja s formalnim nakladnicima u standardizaciji njihove djelatnosti, u smislu optimiziranja procesa dugoročne pohrane.

Sva tri aspekta konteksta pohrane elektroničke građe moraju biti poduprta neprestanim informiranjem, ne samo među stručnjacima koji surađuju, već i u javnosti. Ako se pohrana mrežne građe ili bilo kojeg od opisanih procesa prvi put uvodi, potrebno je uključiti sve stručnjake u organizaciji i informirati ih o tekućim promjenama. Svaka nova organizacijska promjena, ako nije popraćena intenzivnijem informiranjem, može prouzrokovati nesigurnost među stručnjacima, a time i odbijanje njihova sudjelovanja ili čak borbu protiv uvođenja novina. Uključujući ih u procese, rezultati će biti pozitivniji.

Jedno od rješenja moglo bi biti uvođenje informacijskoga portala na kojem se objavljuju i prate sve promjene u organizaciji, gdje su dostupne sve strategije, smjernice, pravilnici, standardi i sve ostalo što je povezano s procesima arhiviranja građe.

Proces A1. Preuzimanje i prihvrat elektroničke mrežne građe

U okviru ovoga procesa odvijaju se tri potprocesa: automatsko prikupljanje mrežne građe, ručno prikupljanje građe od nakladnika ili autora i prikupljanje građe iz web arhiva (vidi **sliku br. 25**).

Automatsko prikupljanje mrežne građe - Potproces A1.1: Na osnovi definiranih kriterija selekcije, u robot unesemo početni popis URL-a, namjestimo opseg i dubinu prikupljanja, maksimalnu veličinu datoteka, vrste datoteka, vrijeme intervala između jednog i drugog prikupljanja ili mrežna mjesta koja treba češće posjećivati, te ga aktiviramo. Rastavljač (*parser*) robota preuzet će URL-e iz popisa u posebnoj bazi i provjeriti jesu li već te mrežne stranice u web arhivu. Ako nisu, preuzet će ih s interneta i nove URL-e uključiti u bazu s popisom ostalih URL-a. Ako pak ti URL-i već jesu u web arhivu, postoje dvije mogućnosti: prva je da provjeri online, a druga da preuzme URL iz arhiva. Ako znamo da se sadržaj određenih URL-a neprestano mijenja, uputimo robota da prvo provjeri online te da preuzme nove i obnovljene URL-e i smjesti ih sa stranicama koje se nisu promijenile i koje su već smještene u web arhivu. Ako se radi o mrežnom mjestu kod kojeg očekujemo manje promjena, onda će robot samo jednom preuzeti URL-e online, a u ostalim će slučajevima te iste URL-e preuzimati iz web arhiva i njegov URL uključiti dalje u popis za sljedeće preuzimanje. U svakom provjeravanju robot pomoću kontrolnoga zbroja MD5 provjerava identičnost posjećenih URL-a.

Specifikacije robota za prikupljanje mreže

Opće specifikacije robota

Implementacija robota mora biti usmjerena prema objektu i, koliko je moguće, pomoću standarda neovisna o platformi. Robot mora:

- biti sposoban da istovremeno preuzima više mrežnih stranica, odnosno da se koristi višeprosesorskom ili višeportnom (*multi-port*) arhitekturom;
- biti sposoban za namještanje brzine prikupljanja; ona se određuje brojem kilobajta prenesenih podataka u jednoj sekundi; ako je brzina veća, mogu se dogoditi zastoji i može doći do blokiranja prijenosa te opterećenje "*hostovog*" poslužitelja;
- razumjeti različite protokole HTTP, FTP, gopher i NNTP; morao bi imati sposobnost odabira ili odstranjivanja protokola, odnosno implementiranja novih protokola;
- razumjeti sve verzije HTML i XML;
- sadržavati dobro poznat SGML DTD ili TEI;
- imati nomadske klijente koji mogu biti namješteni na velikim mrežnim mjestima ili na mrežnim mjestima s malim prometom (*bandwidth lines*);
- imati mehanizam za provjeravanje i predlaganje identičnih mrežnih mjesta koja imaju različita domenska imena;
- održavati bazu podataka o svim mrežnim mjestima i posjećenim dokumentima; u bazi bi se prikupljali sljedeći podatci:
 - o ime mrežnoga mjesta;
 - o broj IP;
 - o dopuštena frekvencija pristupa mrežnom mjestu i parametri za izračun tog podataka;
 - o popis dijela mreže koji robot ne mora posjetiti;
- imati mogućnost uključivanja i isključivanja funkcije *robots.txt*;
- voditi statistiku o pogrešnim i uspješnim prikupljanjima u vremenu;
- kreirati polja za označavanje mrežnog mjesta koje više nije aktivno;
- kreirati polja za podatke o korisničkim ID i lozinkama;
- voditi podatke o različitim lozinkama;
- omogućiti informaciju o broju URL-a u određenom mrežnom mjestu;
- omogućiti informaciju o tipovima poslužitelja;
- u slučaju različitih domenskih imena za isto mrežno mjesto, dati pravi naslov mrežnoga mjesta;
- omogućiti popis svih imena određenoga mrežnoga mjesta;
- omogućiti polje za unos komentara operatera.

Specifikacije robota za preuzimanje mrežnih stranica

Robot bi morao biti sposoban prikupljati mrežne stranice i mrežna mjesta po sljedećim parametrima:

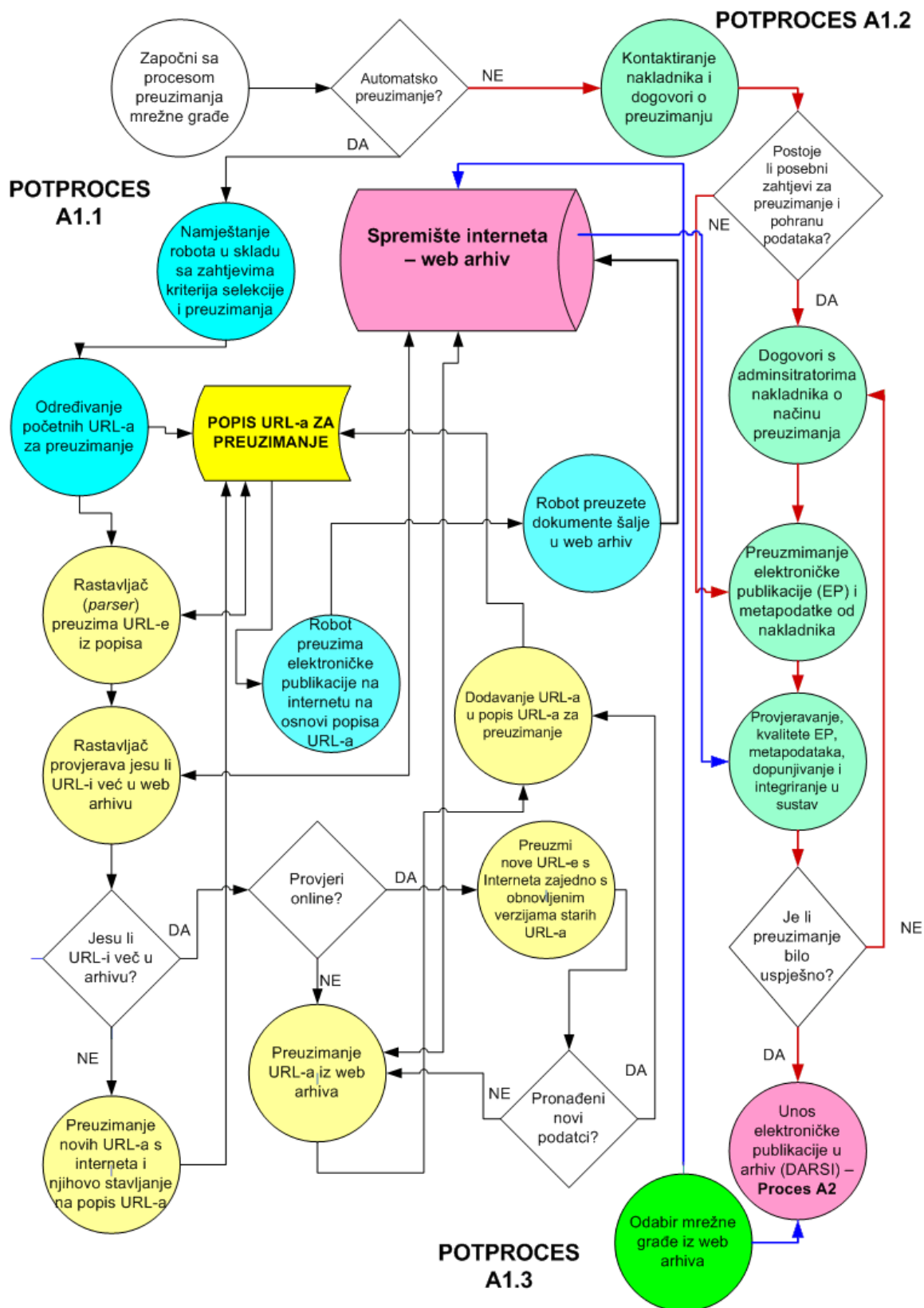
- po popisu mrežnih mjesta na osnovi URL-a,
- na osnovi najviše domene,
- po tipovima datoteka MIME,
- po jeziku tekstova.³⁵⁷

Morala bi postojati mogućnost da se navedeni parametri koriste samostalno ili kombinirano. Osim toga, trebalo bi omogućiti da se za različita mrežna mjesta postavi različita frekvencija preuzimanja.

Robot bi osim toga morao biti sposoban:

- pristupati mrežnim mjestima koja zahtijevaju korisničko ime i lozinku;
- upravljati mrežnim stranicama s okvirima (*frame-based*);
- upravljati zemljovidima;
- prikupiti zaglavlja HTTP (*HTTP-headers*);
- skupljati sve povezane URL-e koji su dio mrežnih stranica (naprimjer, ako se na švedskoj mrežnoj stranici nalazi slika koja je na engleskom poslužitelju, i ona mora biti zahvaćena);
- izdvojiti određen URL koji nakon različitih pokusa nije mogao biti prikupljen;
- prikupiti dokument u cjelini, uključujući sve objekte, bez obzira na poslužitelja na kojem se nalaze;
- imati različite mogućnosti namještanja i u vremenu (naprimjer, dnevno i noćno preuzimanje).

³⁵⁷ Jezična identifikacija mrežnih stranica u eksperimentalnoj je fazi. Na Institutu "Jožef Stefan" pokušali su razviti robot WebBird tako da može prikupljati slovenske mrežne stranice i na osnovi jezika. Dosad je objavljeno više izvještaja stranih istraživanja, među njima: Martins, B. ; Silva, M. Language identification in web pages. // Symposium on Applied Computing: Proceedings of the 2005 ACM symposium on Applied computing, 13.-17. ožujka 1005, Santa Fe, New Mexico. New York : ACM. -URL: http://xldb.di.fc.ul.pt/data/Publications_attach/ngram-article.pdf. (15.12.2006.)



Slika br. 25: Proces A1: Preuzimanje i prihvata mrežne građe

Ručno prikupljanje mrežne građe - Potproces A1.2 - Ručno prikupljanje mrežne građe odvija se u slučajevima kad su potrebna posebna dopuštenja nakladnika ili je publikacija takve prirode (dinamična ili interaktivna) da je za nju potreban poseban postupak prijenosa ili instalacije dodatnoga softvera.

Potproces započinje u trenutku lokalizacije relevantne građe koja je prema kriterijima odabira važna za prikupljanje. Dogovori s nakladnikom ili autorom vode se u dva smjera, s jedne strane u vezi autorskih prava, s druge u vezi dobivanja informacija o tehničkim svojstvima, uvjetima djelovanja građe, važnijim metapodatcima i načinu prijenosa u arhiv. Nakon prihvata građe provjerava se njezina kvaliteta i u slučaju da su opažene pogreške ili da zbog nekog drugog razloga preuzimanje nije bilo uspješno, postupak se ponavlja uz dogovor s administratorima nakladnika ili samim autorom građe. Nakon uspješnog prihvata građa se smješta u radni prostor digitalnoga arhiva.

Prikupljanje građe iz web arhiva - Potproces A1.3 - Pri ručnom prikupljanju potrebno je mnogo više vremena i stručnjaka za lokalizaciju građe, pregovore s vlasnicima autorskih prava i na kraju za obradbu građe. Vjerojatno će doći do situacije kad ćemo utvrditi da postoje publikacije koje su važne korisnicima pa je potrebno olakšati njihovu lokalizaciju s posebnim bibliografskim opisom. To je građa koja više neće biti na mreži ili ona koja je već nekoliko puta promijenjena. Zato ćemo je moći preuzeti iz web arhiva, i to ručno. U ovom slučaju također treba provjeriti kvalitetu i djelovanje građe. Ako se pojave problemi, ne ostaje nam drugo nego da uspostavimo kontakt s autorima ili nakladnicima. Tu ulazimo u potproces A1.2.

Proces A2. Unos elektroničke mrežne građe u arhivski sustav

U nastavku, poslije procesa A1, slijede dva potprocesa koja se odnose na automatsko i ručno prikupljanje građe. (vidi **sliku 26**)

Automatsko prikupljanje građe - Potproces A2.1 - Robot automatski preuzme mrežnu građu i dodaje joj tehničke metapodatke, stvara kontrolni zbroj MD5, vodi i upravlja dupliciranim URL-om. U bazi robota mogu se grupirati URL-i iz jednoga mrežnoga mjesta sa skupnim metapodacima. Kvaliteta se provjerava na uzorcima mrežnih mjesta, a pritom se pazi da poveznice iz različitih razina budu aktivne.

Ručno prikupljanje građe - Potproces A2.2 - Postoji više načina ručnoga prikupljanja građe. Građu možemo prikupljati aktivno, to znači tražiti je, lokalizirati i uspostaviti kontakt s nakladnicima ili autorima, ili pasivno, preko posebnoga sučelja na mrežnim stranicama knjižnice. U **dodatku br. 3** naveden je primjer podataka koji bi mogli biti obuhvaćeni takvim sučeljem.

Koliko će rada zahtijevati ovaj potproces ovisi o vrsti građe. Ako mrežna građa ima svoj ekvivalent u materijalnom mediju, onda je moguće da je već obrađena u bibliografskoj bazi. Optimalno je što više iskoristiti već postojeće bibliografske baze, osim ako se elektronička građa ne želi voditi odvojeno od ostale. U većini nacionalnih knjižnica koje skupljaju elektroničku građu bibliografski opis vodi se zajedno s opisom ostale građe. Ako je elektronička inačica identična s tiskanom, onda se uz bibliografski opis tiskane publikacije može samo dodati napomena da je dostupna i u elektroničkom obliku te navesti adresa pristupa. Ako pak ima dodatne funkcije, onda je poželjno napraviti nov bibliografski zapis.

U nekim sveučilišnim knjižnicama zadovoljni su jednim jedinim bibliografskim opisom elektroničke građe u *Dublin Core*. U nacionalnoj knjižnici to ne bi smjeli dopustiti, jer bi izgubili vrlo bitne informacije o publikaciji. Mnogi zaboravljaju da je dublinski osnovni skup elemenata metapodataka zapravo format namijenjen prije svega izmjeni podataka između različitih ustanova

(naprimjer između knjižnica s jedne strane i arhiva, muzeja ili nakladnika s druge, koji upotrebljavaju vrlo različite standarde opisa) te omogućavanju pristupa toj građi. Treba paziti da se bibliografski opis formalne³⁵⁸ mrežne građe također izvodi stručno, u složenijem formatu, što znači u bilo kojem MARC-u, UNIMARC-u ili odgovarajućem formatu.

Prilikom provjere kvalitete elektroničke građe moramo posebno obratiti pažnju na sljedeće elemente.³⁵⁹

- provjeriti je li građa cjelovito zahvaćena;
- provjeriti je li navigacija po njezinim dijelovima jednaka kao i u originalnom okruženju;
- provjeriti jesu li nadnevak i vrijeme na mrežnim stranicama zamrznuti na vrijeme prikupljanja ili se mijenjaju sa svakim pristupom stranici;
- ako stranice sadrže okvire (*frames*), treba provjeriti jesu li i oni prikupljeni;
- provjeriti na uzorku stranica jesu li tekstovi pravilno prikazani;
- provjeriti na uzorku stranica jesu li slike i ostali elementi pravilno prikupljeni i prikazani;
- ako postoji multimedijски sadržaj, treba provjeriti je li aktivan;
- ako postoje datoteke za preuzimanje, treba provjeriti mogu li se one otvarati;
- ako postoji sučelje za pretraživanje, treba ga provjeriti.

Građu smo preuzeli, smjestili je u radni prostor digitalnoga arhiva. Provjerili smo njezinu kvalitetu, svojstva, metapodatke koje smo dobili od nakladnika ili autora, dodali podatke koji nedostaju, tehničke i strukturne metapodatke. Vrlo je važna informacija o softveru koji podržava njihov pristup. Nakladnici ili autori morali bi sudjelovati u osiguravanju i softvera, ako je on specifičan i nema ga na tržištu. U prošlim poglavljima objasnili smo da je XML odličan kontejner, kako za elektroničku građu, tako i za njezine metapodatke o sadržaju, tehničkim svojstvima i strukturi. Omogućava dugoročnu pohranu i veću fleksibilnost s obzirom na sustav u kojem je pohranjen. Prilikom smještanja u arhivski sustav bilo bi idealno da original odmah pretvorimo u šire prihvaćen³⁶⁰ ili standardni arhivski format³⁶¹ i pohranimo obje inačice,

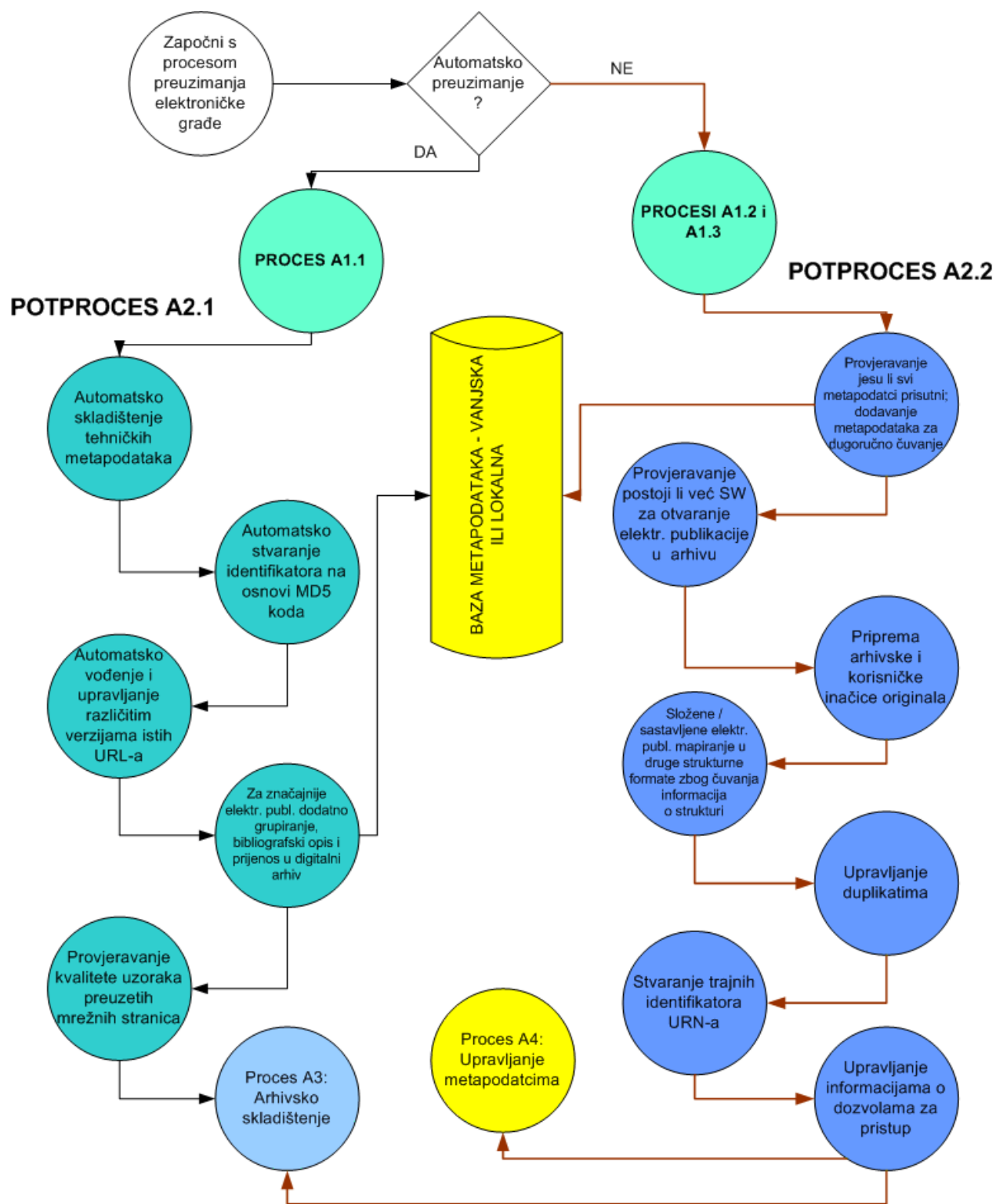
³⁵⁸ U ovom primjeru pod "formalnom" mrežnom građom mislim na građu koja je usporediva s konvencionalnom građom, tj.: knjigama, serijskim publikacijama, člancima itd.

³⁵⁹ Brown, A. Archiving websites : a practical guide for information management professionals. London : Facet, 2006. Str. 71-72.

³⁶⁰ To su formati koji su zbog šire upotrebe manje ugroženi. Naprimjer RTF, PDF ili JPEG. MS Word nikako nije prikladan jer se neprestano razvija.

konvertiranu zbog praćenja razvoja formata, a originalnu u slučaju da se razvoj formata s vremenom pokaže neprimjernim. Budući da su to nezaštićene verzije i često bez ikakvih kompresija te zato zauzimaju mnogo prostora na računalima, bi bilo poželjno izraditi i korisničku inačicu, koja može biti dodatno zaštićena zbog autorskih prava ili prilagođena mrežnom pristupu. Posebnu pažnju moramo posvetiti dodjeljivanju različitih identifikatora svakoj od triju inačica te upravljati s dopuštenjima za pristup toj građi.

³⁶¹ Mnogi se formati razvijaju i svaka verzija ima različite funkcije. Zato su arhivske ustanove između tih verzija odabrali optimalne formate. Takav je naprimjer PDF/A, koji je proglašen standardnim formatom za tekstualne datoteke. Za slike je optimalan format TIFF jer ne sadrži kompresije.



Slika br. 26: Proces A2: Unos elektroničke građe u arhivski sustav

Proces A3. Arhivsko skladištenje mrežne građe

Ovaj proces obuhvaća tri potprocesa, koji se odnose na način unosa građe u arhivski sustav: skladištenje automatski prikupljene građe (A3.1), skladištenje mrežne građe dobivene ručnim unosom (A3.2) i skladištenje građe dobivene na osnovi odabira iz web arhiva (A3.3). (vidi **sliku br. 27**)

Skladištenje automatski prikupljene mrežne građe - Potproces A3.1 - Robot automatski prikupljenu mrežnu građu mapira u XML kontejner i smješta je u web arhiv. Vrlo je važno da prilikom procesa prikupljanja robot apsolutne adrese domena pretvori u relativne, kako se korisnicima ne bi dogodilo da ih prilikom pretraživanja sustav baci u živo okruženje interneta, umjesto na URL u web arhivu. Mrežne stranice moraju biti smještene tako da se mogu lokalizirati cijela mrežna mjesta, različite inačice mrežnoga mjesta u vremenu i pojedinačni URL-i.

Specifikacije robota za pohranu mrežnih stranica

Način pohrane ovisi o metodi i učestalosti prikupljanja te o zahtjevu za pristup web arhivu. Postoje sljedeći mogući parametri:

- mrežna mjesta i dokumenti prikupljaju se u cjelini i svaka zbirka koja čini cjelinu posebno je označena;
- smještanje mrežnih stranica mora biti tako organizirano da dozivanje originalnog mrežnog mjesta bude jednostavno;
- kod svih objekata moraju biti pohranjeni i sadržaj i zaglavlje (*headers*);
- objekti moraju biti smješteni s jedinstvenim identifikatorom, imenom koje ne sadrži "nepouzdana" oznake;
- zajedno s podacima morao bi postojati popis URL-a s odgovarajućim imenima datoteka koje su smještene u web arhivu;
- mora postojati mogućnost dodavanja vremenskoga žiga;

- svi parametri koji određuju zbirku morali bi se posebno pohraniti;
- URL-i se ne bi smjeli duplicirati u web arhivu.

Specifikacije robota za administriranje/vođenje podataka

- Robot bi morao imati odgovarajuće sučelje za administriranje podataka. Preporučljiv je GUI jer omogućava konfiguraciju parametara robota.
- Zajedno s administriranjem baze podataka trebalo bi biti moguće pratiti i djelovanje robota.
- U sučelju GUI trebalo bi biti moguće upravljati s pogrešnim URL-ima (*error 404*, itd). GUI bi trebao omogućiti automatsko upravljanje s tim pogreškama.
- Robot bi automatski morao obavještavati jesu li pravila *robots.txt* promijenjena u određenom mrežnom mjestu.
- Morao bi omogućiti uvid u popis URL-a koji čekaju na preuzimanje.
- Morala bi postojati mogućnost brisanja URL-a s popisa za preuzimanje, individualno ili na osnovi određenoga algoritma.
- Trebalo bi biti moguće zaustaviti i ponovno pokrenuti robot, a da se pritom ne prouzroči nikakva šteta.
- Trebala bi postojati mogućnost da se robot postavi u pasivno stanje (*standby*) u kojem bi se zaustavila sva aktivnost.

Specifikacije robota za vođenje statistike

Robot bi morao sadržavati modul za statistiku koji bi davao barem sljedeće podatke:

- broj i postotak procesiranih URL-a,
- broj i postotak vraćenih kodova,
- broj i postotak novih izvađenih URL-a iz podataka,
- novi URL-i koji su sortirani po mrežnom mjestu,
- uporaba *bandwidth*,
- statistika o postojećim metapodacima.

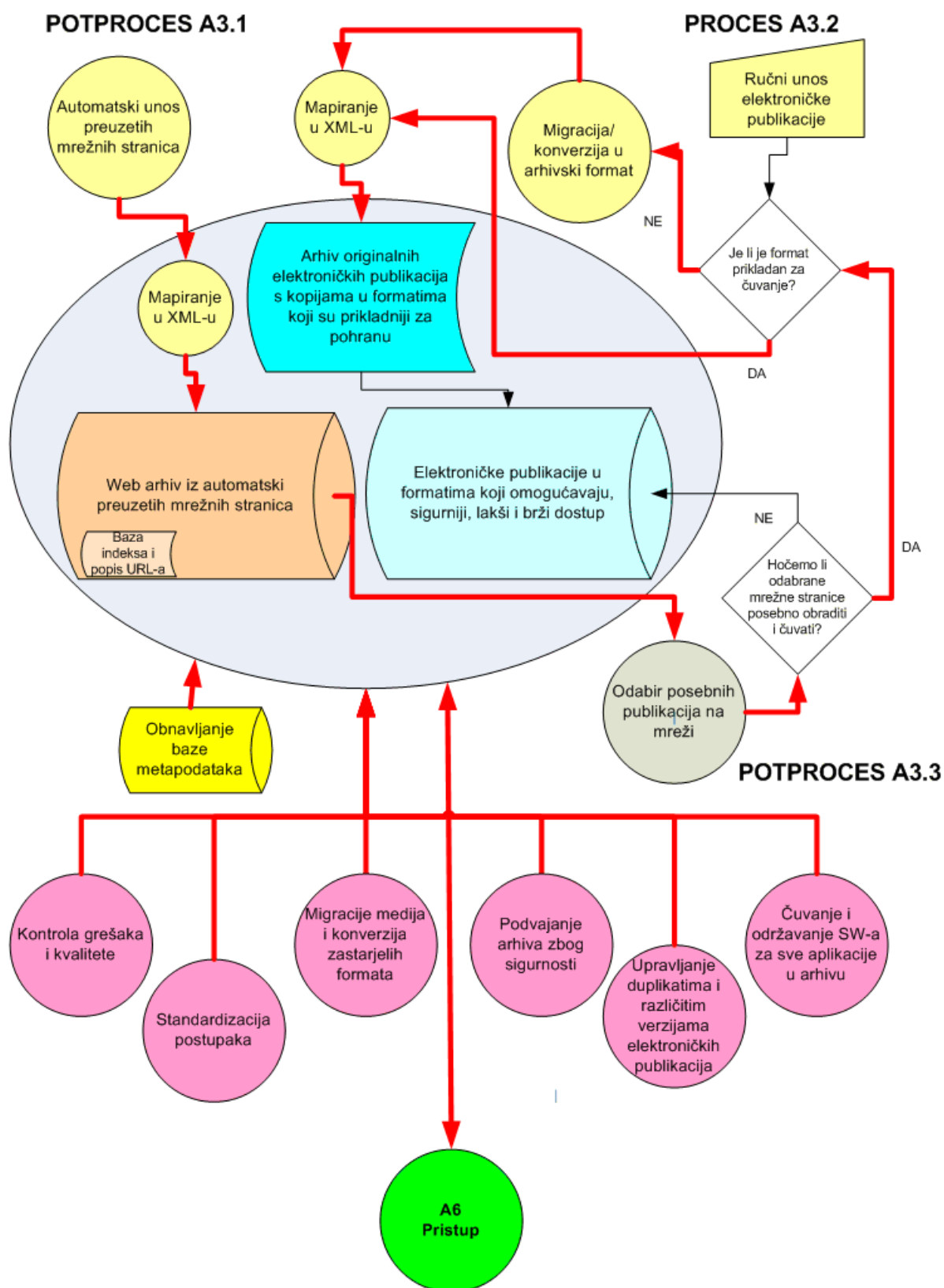
Skladištenje mrežne građe dobivene ručnim unosom - Potproces A3.2 - U prošlom procesu vidjeli smo da je potrebno voditi računa o tome da se datoteke sačuvaju u tri različita formata:

originalnom, konvertiranom u općeprihvaćen format te korisničkom format. Naravno, može se dogoditi da je original već u općeprihvaćenom formatu ili da su sve tri datoteke u korisničkom formatu. Ako taj format ne odgovara standardima za dugoročnu pohranu, onda bi od nakladnika ili autora morali tražiti bolju inačicu. XML je rješenje ne samo za arhivsku kopiju, već i za ostale. Međutim, barem bi original morao biti mapiran u XML. METS ili MODS već su pripremljeni formati za elektroničku građu s opisom u MARC 21. Portugalci su preradili METS za potrebe bibliografskoga opisa u UNIMARC-u.³⁶²

Skladištenje građe dobivene na osnovi odabira iz web arhiva - Potproces A3.3 - Treći potproces obuhvaća građu koju bi automatski prikupio robot, a zatim bi je bibliotekari odvojeno uključili u digitalni arhiv. Tu građu tretiramo kao i ručno preuzetu građu i zato je uključujemo u A.3.2 tako što analiziramo njezinu kvalitetu i odlučujemo o arhivskom i korisničkom formatu.

Kod sva tri potprocesa vrlo je važno da se obnavljaju sve baze metapodatka i da arhivski sustav omogućava neprestanu kontrolu pogrešaka i kvalitete, da omogućava izvođenje migracije medija i konverzije zastarjelih formata. Arhivski sustav mora biti sposoban upravljati duplikatima i različitim verzijama iste građe i mora zbog sigurnosti osigurati svoju zrcalnu kopiju. Ako nije moguće istovremeno održavanje softvera za pristup svim aplikacijama u arhivu, mora biti predviđena alternativa kojom bi se omogućio pristup tim aplikacijama. Svi arhivski postupci u okviru ovoga procesa moraju biti usklađeni i temeljiti se na važećim standardima.

³⁶²Borbinha, J. ; Freire, N. ; Neves, J. BND: the architecture of a National Digital Library. Paper presented at Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries, 2004. 7-11 June 2004, str. 21- 22, DOI:10.1109/JCDL.2004.1336090. URL: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=1336090&isnumber=29473. (15.12.2006.)



Slika br. 27: Proces A3: Arhivsko skladištenje mrežne građe

Proces A4. Upravljanje metapodacima

U arhivskom sustavu imamo dvije vrste repozitorija: repozitorij web arhiva u koji se pohranjuje automatski prikupljena mrežna građa i repozitorij digitalnog arhiva u kojem su publikacije obrađene jednim ili zajedničkim bibliografskim opisom (vidi **sliku br. 28**). Da bismo našli građu u web arhivu, ona se mora indeksirati. Ako želimo pretraživati po njegovu punom tekstu, onda se indeksiraju ne samo URL adrese i tehnički i administrativni podatci, kao što su MIME tipovi i veličina datoteka, vrijeme prikupljanja itd., već i sadržaji mrežnih stranica. Na taj će način biti omogućen sličan pristup kao i u mrežnom živom okruženju.

S druge strane, građa u digitalnom arhivu tretirana se kao tradicionalna građa: ona je samostalna jedinica i ima svoj bibliografski opis. Međutim, tu treba razlikovati dodatne vrste metapodataka, koje tradicionalna građa nema, a to su metapodatci za dugoročnu pohranu, administrativni i tehnički metapodatci. Ti se metapodatci mogu unijeti u strukturu XML i na taj način omogućiti pretraživanje po arhivskoj bazi.

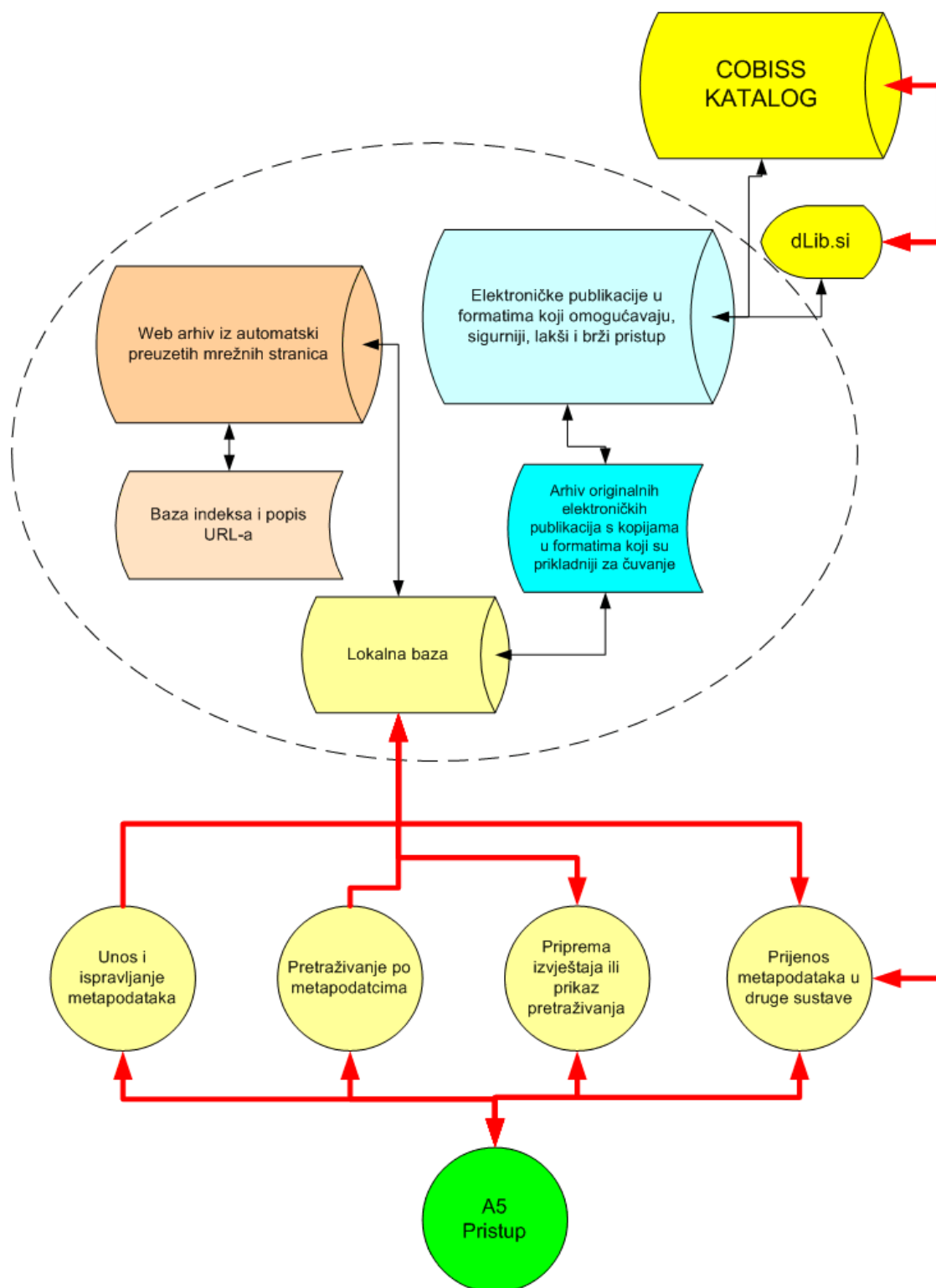
Ako ne koristimo XML, onda bismo morali sve te metapodatke unijeti u neku bazu po kojoj bismo mogli pretraživati. Postoje dvije mogućnosti, prva je da te metapodatke uključimo u već postojeću glavnu bibliografsku bazu tako što ćemo je proširiti za dodatna polja. Drugo je rješenje da te podatke vodimo odvojeno u posebnoj bazi u okviru digitalnoga arhiva, u kojoj je uspostavljena relacija putem identifikatora sa zapisom u glavnoj bazi. Bibliografska bi baza imala potpun bibliografski opis po složenijem formatu (MARC/UNIMARC), a lokalna baza digitalnoga arhiva koristila bi dublinski osnovni skup elemenata metapodataka, kojima bi se dodali metapodatci za dugoročnu pohranu, kao i administrativne i tehničke metapodatke u slučaju da oni već nisu obuhvaćeni samim dokumentom. Koja će se baza prva puniti, ovisi od tomu ima li elektronička publikacija svoju materijalnu (analognu) inačicu u glavnoj bibliografskoj bazi. U tom bi se slučaju dodali metapodatci u glavnoj bazi, odakle bi se prenijeli u lokalnu bazu, gdje bi se dodali podatci o dogovorima o autorskim pravima, način pridobivanja publikacije, svojstava, način dostupa itd. Taj bi postupak morao biti automatiziran tako da knjižničari ne gube vrijeme na otvaranje različitih baza. Noviji softveri za digitalne knjižnice, kao što su Fedora iz sveučilišta

u Virginiji, SAD, DSpace iz MIT-a kod Bostona, DigiTool iz ExLibrisa i ostali, omogućavaju udruživanje različitih mogućnosti u samom knjižničnom sustavu. U hrvatskoj Nacionalnoj i sveučilišnoj knjižnici također su našli dobro rješenje za taj proces.

U slovenskoj Narodnoj i sveučilišnoj knjižnici već imamo OPAC, bibliografski katalog COBIB u okviru sustava COBISS, iz kojega se izvlače bibliografski metapodaci u formatu ISO. Pretraživanje po različitim zbirkama u portalu dLib.si omogućeno je protokolom OAI-PMH. Inače su elektroničke publikacije dostupne preko samog kataloga. Budući da je COBIB vrlo nefleksibilan, teško je prilagoditi strukturu zapisa pa smo planirali izgraditi posebnu bazu metapodataka za dugoročnu pohranu, u koju unosimo informacije lokalnoga značenja.

Lokalna i glavna bibliografska baza moraju omogućiti unos i ispravljanje metapodataka, pretraživanje po svim metapodacima, pripremu izvještaja ili prikaz pretraživanja te prijenos metapodataka u druge sustave.

PROCES A4



Slika br. 28: Proces A4: Upravljanje metapodacima

Proces A5: Korisnički pristup

Svrha korisničkoga pristupa je omogućiti korisnicima da pretražuju i pronađu željenu građu te pristupe njezinom sadržaju (vidi **sliku br. 29**). Pritom je prisutno više aspekata. Kao prvo, moramo definirati koje ćemo sve mogućnosti pretraživanja i prikazivanja podataka pružiti korisnicima. U nastavku slijedi prijedlog definicije funkcionalnosti korisničkoga sučelja:

Pretraživanje mora biti omogućeno kako u web arhivu, tako i u bibliografskoj bazi.

Pretraživanje po web arhivu moralo bi biti slično kao i kod svjetskih pretraživača, s dodatnom mogućnošću pretraživanja i u vremenu. Sučelje bi moralo omogućiti sljedeće funkcije:

- traženje po punom tekstu ili indeksima; rezultati pretraživanja prikazali bi se na principu frekventnosti pojave traženoga pojma;
- traženje po nadnevcima;
- traženje po URL-ima.

Pretraživanje bi se izvodilo:

- pomoću Boolovih operatera (i, ili, ne)
- distribuiranim pretraživanjem (kombinacija više mogućnosti)
- na osnovi upisanih riječi ili fraza
- pomoću desnoga kraćenja riječi.

Specifikacije sučelja za pristup web arhivu

Glavni cilj projekta NWA bio je izgradnja sučelja za pristup web arhivu. Tako se cijeli projekt temeljio na detaljnoj specifikaciji sučelja.³⁶³ Rješenja iz NWA mogu se koristiti i u našem primjeru. U nastavku su prikazani osnovni parametri za izgradnju sučelja za pristup NWA arhivu:

1. **Autentičnost** - dokumenti se pohranjuju bez promjena i identični su s originalom; jedno od rješenja je upotreba kontrolnoga zbroja MD5 (*checksum*) i stvaranje vremenskih žigova;
2. **Format** - sučelje bi moralo poduprijeti sve tipove MIME datoteka;

³⁶³Nordic Web Archive. Introduction. Appendix II - Specification for an archiving robot. Pp. 15-18 URL: <http://www.lib.helsinki.fi/tietolinja/0100/nwa.pdf> (15.12.2006.)

3. **Identifikacija** - svi prisutni identifikatori u dokumentu moraju se indeksirati, međutim oni ne bi smjeli zamijeniti ID u arhivu jer se jednaki bibliografski identifikatori dodjeljuju i različitim izdanjima elektroničke publikacije (naprimjer članci imaju isti ISSN broj i u dokumentu identifikator može označivati samo referencu za drugi dokument); u NWA su odlučili kao ID koristiti kontrolni zbroj MD5 koji bi se mogao proširiti u URN dokumente;
4. **Pristup** (navigacija, indeksiranje, traženje) - budući da pristup ovisi o zakonodavstvu u različitim zemljama, sučelje za pristup mora imati mogućnost za ograničavanje ili prilagođavanje potrebama različitih okruženja;
 - a. **Indeksiranje** - započinje s analizom formata datoteka; pritom će se neki formati, naprimjer slike, ograničiti na automatsko računalno generiranje metapodataka; kod tekstova će se primijeniti lingvističke metode za indeksiranje; indeksi se pohranjuju u posebnoj bazi;
 - b. **Traženje i pretraživanje** - trebalo bi pokazati bibliografsku referencu do elektroničke građe; na projektu NWA predlagali su sljedeće funkcije pretraživanja:
 - traženje po cijelom tekstu - napredno i opće traženje,
 - traženje po metapodacima koje generira aplikacija za preuzimanje, naprimjer ID i vremenski žig,
 - traženje po domenama/geografskim lokacijama,
 - traženje po organizacijama,
 - traženje na osnovi fraza ili ključnih riječi,
 - traženje po formatima ili drugim parametrima;
 - c. **Navigacija** - šetanje po mrežnoj stranici i prijelaz iz jedne poveznice u drugu - poveznice moraju djelovati u samom web arhivu; sustav bi morao omogućiti:
 - navigaciju po web arhivu,
 - navigaciju po poveznicama jednako kao u originalnom okruženju u oba smjera,
 - navigaciju s praćenjem ulaznih poveznica (dokumenata koji su povezani s tom stranicom) i mogućnost povezivanja sa živim okruženjem,

- navigaciju po vremenu tako da je moguć uvid u različita izdanja jedne mrežne stranice u vremenu;

d. **Korisničko sučelje u web arhivu** - trebalo bi biti mrežno i moralo bi imati mogućnost upotrebe sa standardnim pretraživačima, kao što su *Netscape*, *Internet Explorer*, *Mozilla Firefox* i sl.; bog različitih formata datoteka potrebni su različiti dodaci (*plug-ins*).

Pretraživanje po digitalnom arhivu

Ako već postoji OPAC sustav, onda su parametri za jedinice građe čiji bibliografski opis je dio tog sustava već određeni. U suprotnom, ako omogućavamo pristup iz druge bibliografske baze, onda treba iznova definirati sve važnije parametre. Sučelje bi moralo omogućiti sljedeće funkcionalne zahtjeve:

- traženje po određenim poljima i bilo gdje
- traženje s Boolovim operaterima
- distribuirano pretraživanje
- traženje na osnovi upisanih riječi ili fraza
- traženje pomoću desnog kraćenja riječi.

Osim toga, sustav bi morao omogućiti sljedeće:

- mogućnost pohrane prošlih pretraživanja i njihovu međusobnu kombinaciju
- postojanje ugrađenoga rječnika ili indeksa za traženje po poljima
- u svakom bi se pretraživanju morao pokazati broj dobivenih rezultata
- morala bi postojati mogućnost pregledavanja i odabira jedinica građe u okviru jednoga pretraživanja
- morala bi postojati mogućnost ograničavanja traženja s nadnevkom izdanja publikacije.

Rezultati bi morali imati mogućnost prikazivanja po godini publikacije, po autoru, po naslovu ili po viru.

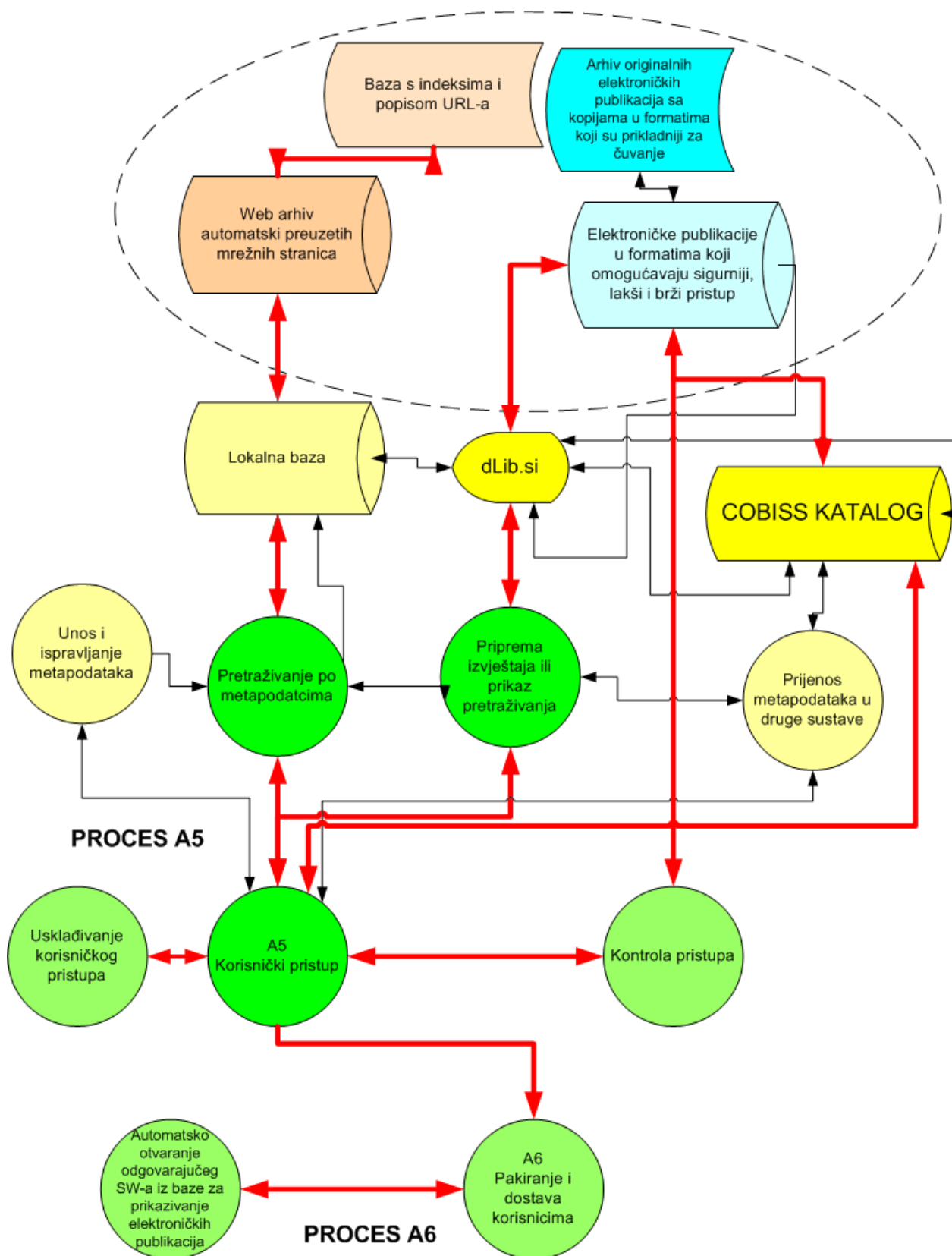
Proces A6: Pakiranje i dostava

Kada korisnik nađe traženu građu, u istom trenutku dolazi do provjeravanja je li mu je dopušten pristup njezinu sadržaju. Obično bi sva elektronička građa morala biti dostupna barem u lokaciji same knjižnice. ako postoji mogućnost mrežnoga pristupa u digitalni arhiv, onda je potreban složeniji sustav koji bi na osnovi IP brojeva ili lozinke kontrolirao pristupe, odnosno ograničavao bi pristup onoj građi koja je autorski zaštićena.

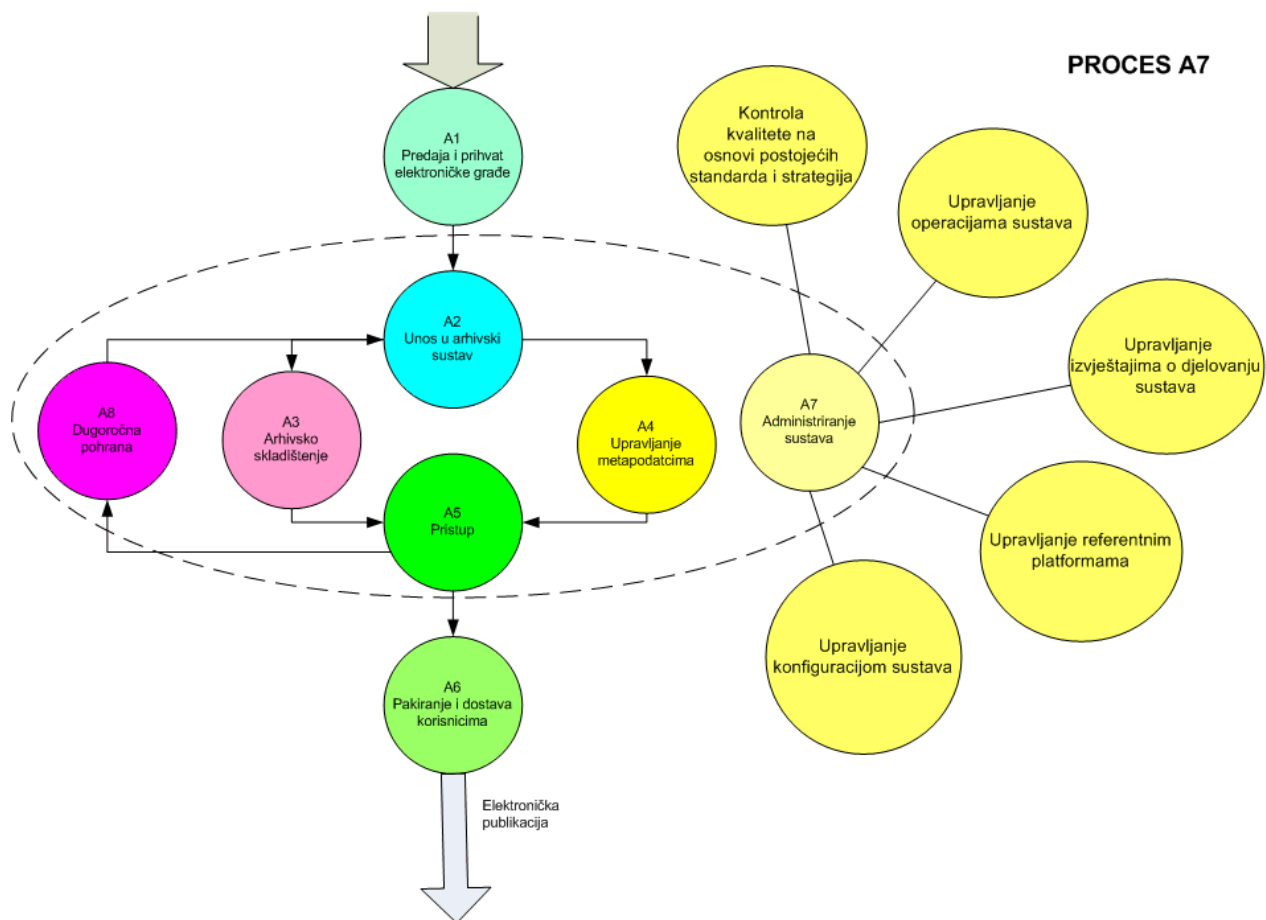
Osim spomenutog, u procesu pristupanja građi moramo uvijek aktivirati ili pokrenuti primjeren softver. Danas je to još uvijek jednostavno jer standardne aplikacije otvaraju većinu producirane građe. Problem bi mogao nastati za deset godina.³⁶⁴ Zato je važno da se u okviru procesa za dugoročnu pohranu prate promjene i izvode konverzije datoteka u pravo vrijeme. U okviru baze PRONOM vodi se popis svih mogućih računalnih formata i mogućih programa kojima se datoteke mogu pročitati.

U trenutku kada je korisnik dobio odobrenje da pristupi elektroničkom sadržaju, mora se aktivirati softver i sadržaj "spakirati" na takav način da ga korisnik može pročitati i razumjeti. Korisniku bi se usto mogle pružiti i kontekstualne informacije u vezi s građom.

³⁶⁴ Mnogi se još sjećamo starog programa *WordStara*, koji se prije više od 10 godina prestao razvijati. Datoteke napisane u tom programu mogle su se konvertirati i pročitati pomoću *WordPerfecta*. Danas je ta aplikacija već odavno zaboravljena i kad bi knjižnica pohranjivala građu napisanu njome, morala bi osigurati i način pristupa. Isti je slučaj i s programima koji su napisani za *Sinclair* ili *Commodore*.



Slika br. 29: Proces A5: Korisnički pristup i Proces A6: Pakiranje i dostava

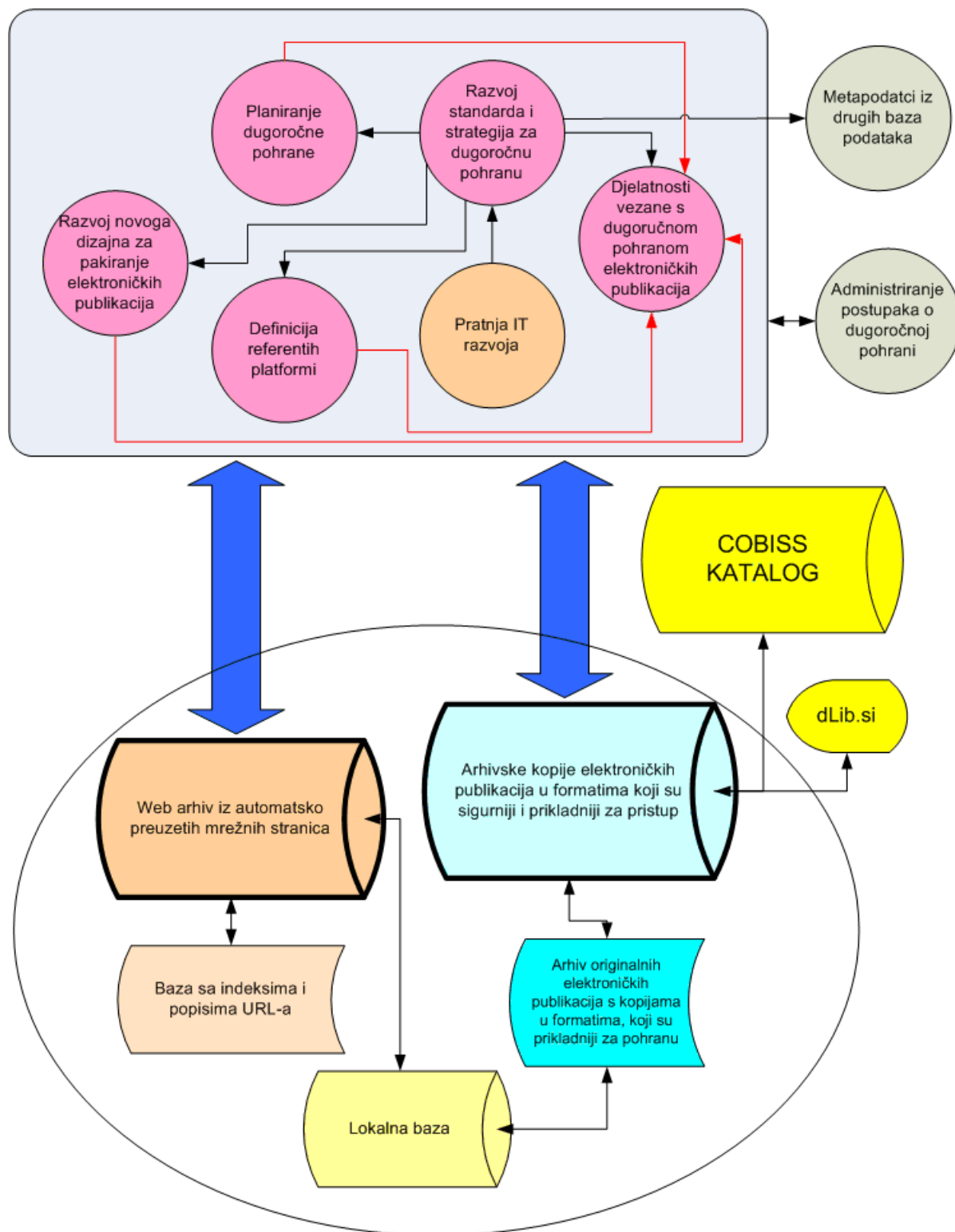


Slika br. 30: Proces A7: Administriranje arhivskoga sustava

Proces A7: Administriranje arhivskoga sustava

Administriranje arhivskog sustava odnosi se na upravljanje svim informacijsko-tehnološkim procesima u samom arhivskom sustavu. (vidi **sliku br. 30**). Taj proces obuhvaća kontrolu kvalitete djelovanja sustava na osnovi postojećih standarda i strategija, upravljanje operacijskim sustavima u arhivskom sustavu, upravljanje izvještajima o djelovanju sustava, upravljanje referentnim platformama i konfiguracijom sustava itd.

PROCES A8



Slika br. 31: Proces A8: Dugoročna pohrana

Proces A8: Dugoročna pohrana mrežne građe

To je proces koji podrazumijeva neprestano planiranje strategija dugoročne pohrane (migracije, konverzije i ostalo), praćenje domaćih i međunarodnih standarda i strategija za dugoročnu pohranu, razvoj novih dizajna za pakiranje elektroničke i mrežne građe, organizaciju svih djelatnosti u vezi sa dugoročnom pohranom u arhivskom sustavu, definiciju potrebnih metapodataka za dugoročnu pohranu, obnavljanje svih informacija i metapodataka, praćenje kvalitete građe, praćenje i predviđanje razvoja informacijske tehnologije i softvera, definiciju novih referentnih platformi, usklađivanje svih postupaka na području dugoročne pohrane sa *back-up* lokacijom i slično (**slika br. 31**). Uspješnost ovoga procesa ovisi o dobro definiranoj politici duge pohrane elektroničke građe, dovoljnom broju stručnjaka koji mogu sudjelovati u tom procesu te ga pratiti i surađivati sa sličnim nacionalnim i međunarodnim institucijama koje imaju funkciju depozitarne ustanove za elektroničku građu. Budući da se područje tek razvija i da su standardizacije u tijeku, ne možemo govoriti o standardnoj praksi ili optimalnim rješenjima. Prošlo je samo 15 godina od pojave interneta, a dugoročna pohrana će tek za 50 godina pokazati koja su rješenja bila optimalna. Najviše što u ovom trenutku možemo učiniti jest pokušati pohraniti što više informacija o digitalnim objektima.

12. ISKUSTVA NA PROJEKTIMA NUK-a

Krajem 2002. godine NUK je u suradnji sa IJS-om započeo projekt "Metodologija skupljanja i arhiviranja slovenskih elektroničkih publikacija na mreži". To je bilo vrijeme zatišja, dvije godine nakon završetka projekta NEDLIB i prve implementacije modela OAIS za potrebe knjižnica, kad je izvještaj *Trusted digital repositories* bio tek objavljen, AACR2 je donio novu podjelu građe, novozelandska je knjižnica tek izradila svoj izbor metapodataka za dugoročnu u pohranu, a robot NEDLIB Harvester, unatoč pogreškama i problemima, bio je najviše dostignuće alata za automatsko prikupljanje mreže. Glavni razlog za suradnju s istraživačkim institutom IJS bio je taj da su se oboje ustanove javile na natječaj za financiranje projekata u okviru Ciljnih istraživačkih programa "Konkurentnost Slovenije 2001-2006". Njihovi su se projekti sadržajno dopunjavali. IJS je prijavio projekt izrade robota za prikupljanje građe, a NUK je prijavio projekt studije izvodivosti za prikupljanje slovenskih publikacija na mreži. Ministarstvo za kulturu, Ministarstvo za obrazovanje, znanost i sport te Ministarstvo za informacijsko društvo, koji su tada financirali projekte, predložili su da zajedno udružimo snage i znanje te razvijemo zajednički projekt. Projekt je formalno trajao do kraja listopada 2004. godine.

Na IJS-u su 2003. godine utvrdili da opseg slovenske mreže čini oko 10 milijuna mrežnih stranica s tekstem, što je tada iznosilo oko 500 GB podataka.³⁶⁵ Te su se mrežne stranice nalazile u okviru protokola HTTP. Krajem lipnja 2004. godine u ARNES-u je bilo registrirano 17.500 domena iz akademskoga i javnoga sektora. Ti su brojevi omogućavali kontrolu i jednostavno prikupljanje slovenske mreže. Problem su bile dinamične stranice u JavaScriptu i Flashu, koje noviji roboti tipa Heritrix još nisu sposobni prepoznati.

Na projektu smo polazili od tradicionalnih i formalnih koncepata publikacija i odabrali smo četiri vrste koje se razlikuju po obliku i funkciji: naslovnice, knjige, serijske publikacije (neomeđenu

³⁶⁵ Danas je taj broj nešto veći. Opće pretraživanje u Googlu daje rezultat 24.400.000 mrežnih stranica s domenom na *.si, što je 12 puta više nego li 2003. godine (2.040.000 stranica na *.si). U okviru te domene moguće je naći ukupno 1.200.000 datoteka u PDF-u, 89.000 datoteka u MS Wordu, 9.800 na *.rtf, 11.700 na *.xls.

građu) i kartografsku građu. Među prve uvrstili smo svu građu koja se nije mogla svrstati u ostale tri kategorije. Elektroničke knjige su po formi i funkciji odgovarale tiskanim knjigama,³⁶⁶ jednako je bilo i s neomeđenom građom,³⁶⁷ dok je kartografska građa³⁶⁸ na neki način rješavala i problematiku složene slikovne građe.

Veći rezultati na projektu bili su sljedeći:

1. Odredili smo osnovnu strategiju prikupljanja građe - odredili smo kriterije selekcije za sve četiri vrste mrežne građe.
2. Pojednostavili smo bibliografski opis u UNIMARC-u za potrebe svih četiriju vrsta građe (vidi **dodatak br. 1**), koji bi služio kao osnova za lokalnu bazu.
3. Odredili smo metapodatke za dugoročnu pohranu građe, koji su objavljeni u članku u *Knjižnici*.³⁶⁹
4. Da bismo unijeli te podatke, dodali smo nova polja u IX. bloku COMARC-a.³⁷⁰ (**tabela br. 8**)

³⁶⁶ Najveći nakladnici elektroničkih knjiga u Sloveniji su javni zavodi, prije svega sveučilišta i javna uprava. U tu se skupinu ubrajaju normativni akti, pravilnici, godišnji i ostali izvještaji, priručnici, upute, programi, šifranti, definicije, metodologije, istraživački radovi itd.

³⁶⁷ U Sloveniji su u rujnu 2003. godine bila registrirana 184 elektronička nakladnika, koji su objavili 200 elektroničkih serijskih publikacija s odgovarajućim ISSN. Većina su bili pojedinci (53), poduzeća (46), sportska društva i udruge (25). Najviše njih objavljivalo je elektroničke novine (82). Vidi: Prilog 2: Kodrič, N. Analiza slovenskih elektronskih serijskih publikacij in njihovih založnikov. // Kavčič-Čolić, A. ... [e tal.]: *Metodologija zbiranja in arhiviranja slovenskih elektronskih publikacij na medmrežju* / [nositeljica projekta Kodrič-Dačić, E. ; projektna skupina Kavčič-Čolić, A. (rukovodilac projekta) ... et al.]. - Ljubljana : Narodna in univerzitetna knjižnica, 2004. 105 str. (Ciljni raziskovalni programi).

³⁶⁸ Geodetska uprava Republike Slovenije zadužena je za uspostavljenje, održavanje i vođenje državnog topografsko-kartografskog sustava u grafičkom, analognom i elektroničkom obliku. Vidi: Šolar, R. Digitalno kartografsko gradivo, nov izziv kartografskih zbirk // *Knjižnica* 47, 4(2003), str. 7-22.

³⁶⁹ Kavčič-Čolić, A. Metapodatki za trajno ohranjanje elektronskih virov. // *Knjižnica* 48, 4(2004), str. 97-119.

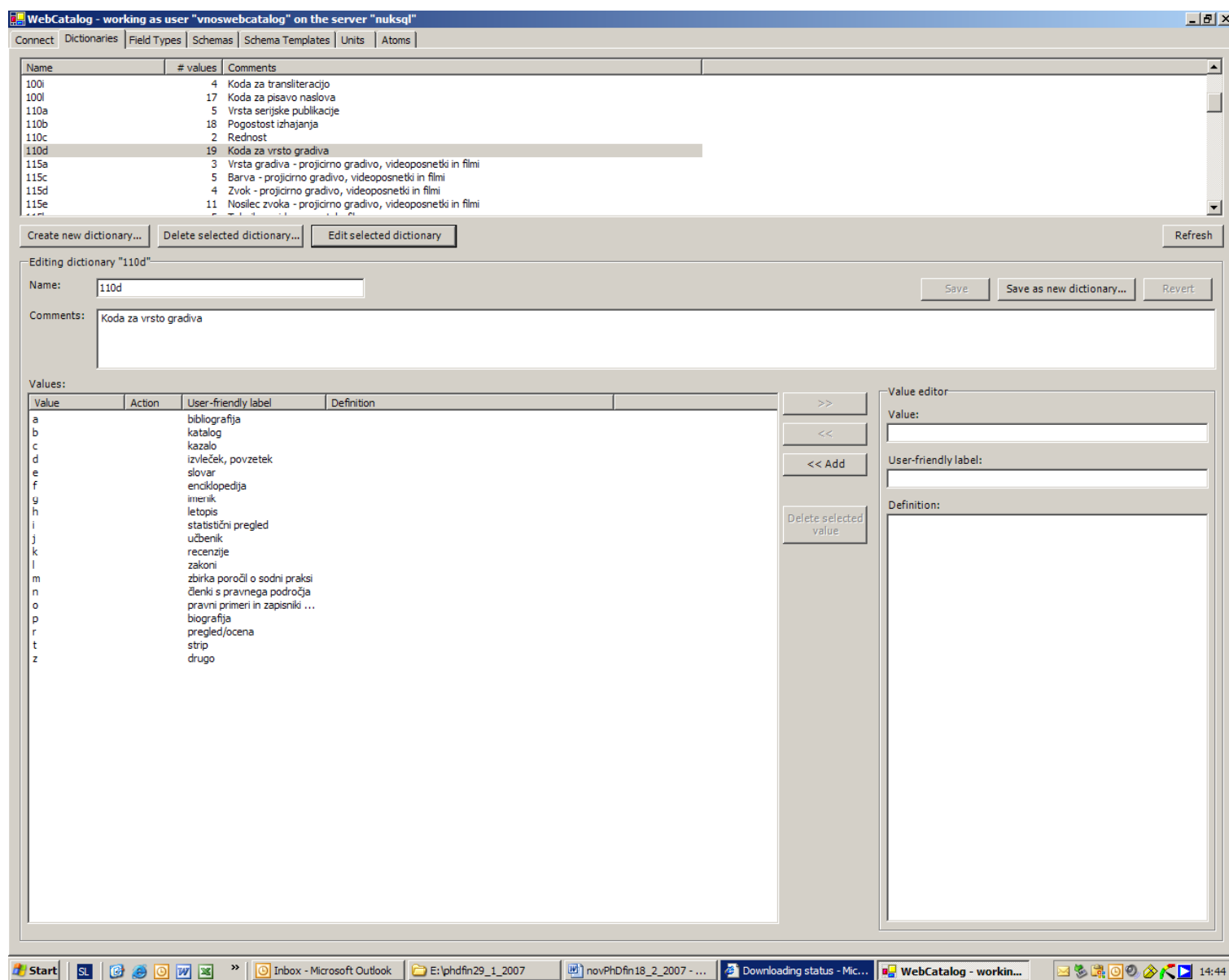
³⁷⁰ COMARC je slovenski UNIMARC.

Tabela br. 8: Dodatna polja u COMARC-u za metapodatke za dugoročnu pohranu

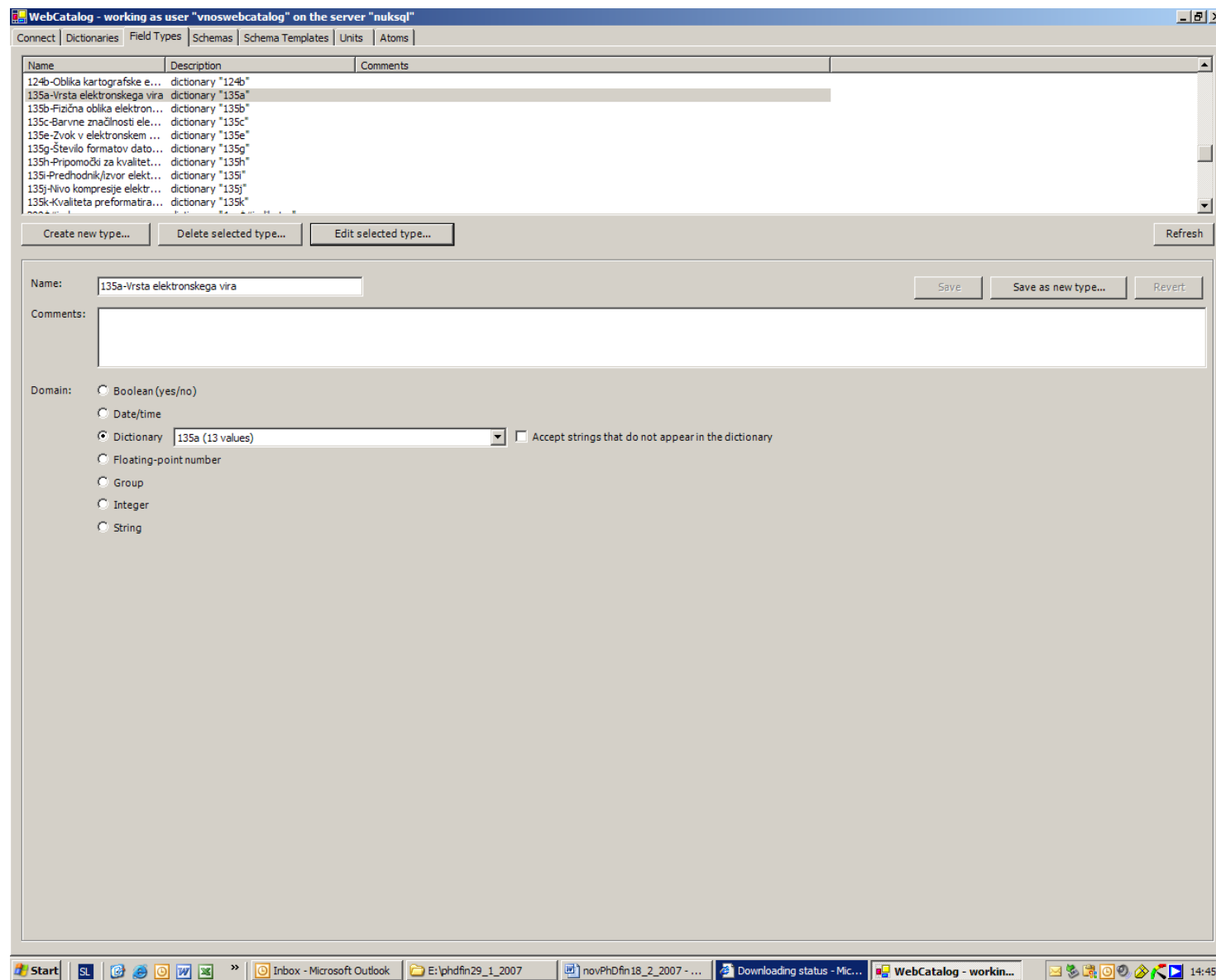
OZNAKA POLJA I POTPOLJA	IME POLJA I POTPOLJA	PONOVLJIVOST POLJA I POTPOLJA P=ponovljivo N=neponovljivo
950	Podatci o promjenama elektroničke građe	P
\$a	Datum promjena (LLLLMMDD)	N
\$b	Opis promjena	N
\$c	Autor promjene	N
951	Podatci o autorskim pravima	P
\$a	Nositelji autorskih prava	N
\$b	Broj ugovora	N
\$c	Vrijeme zaštite	N
\$d	Uvjeti pristupa	N
\$e	Dopuštene intervencije u elektroničkoj građi	N
\$f	Osobe kojima je pristup dopušten	N
\$g	Ostali važni podatci koji nisu navedeni u potpoljima od \$a-\$g	
952	Informacije o autentičnosti (<i>Fixity</i>)	P
\$a	Kontrolni zbroj (<i>Checksum</i>)	N
\$d	Ostalo	P
953	Upute za upotrebu	P
\$a	Opis	N
954	Kompajleri i interpreteri za različite programske jezike	P
\$a	Ime	
\$b	Verzija	N
\$c	Upute	
955	Ostali podatci o pohrani	P
\$a	Opis	N

Ta su polja dopunjavala informacije iz polja 856 (vidi **dodatak br. 1**)

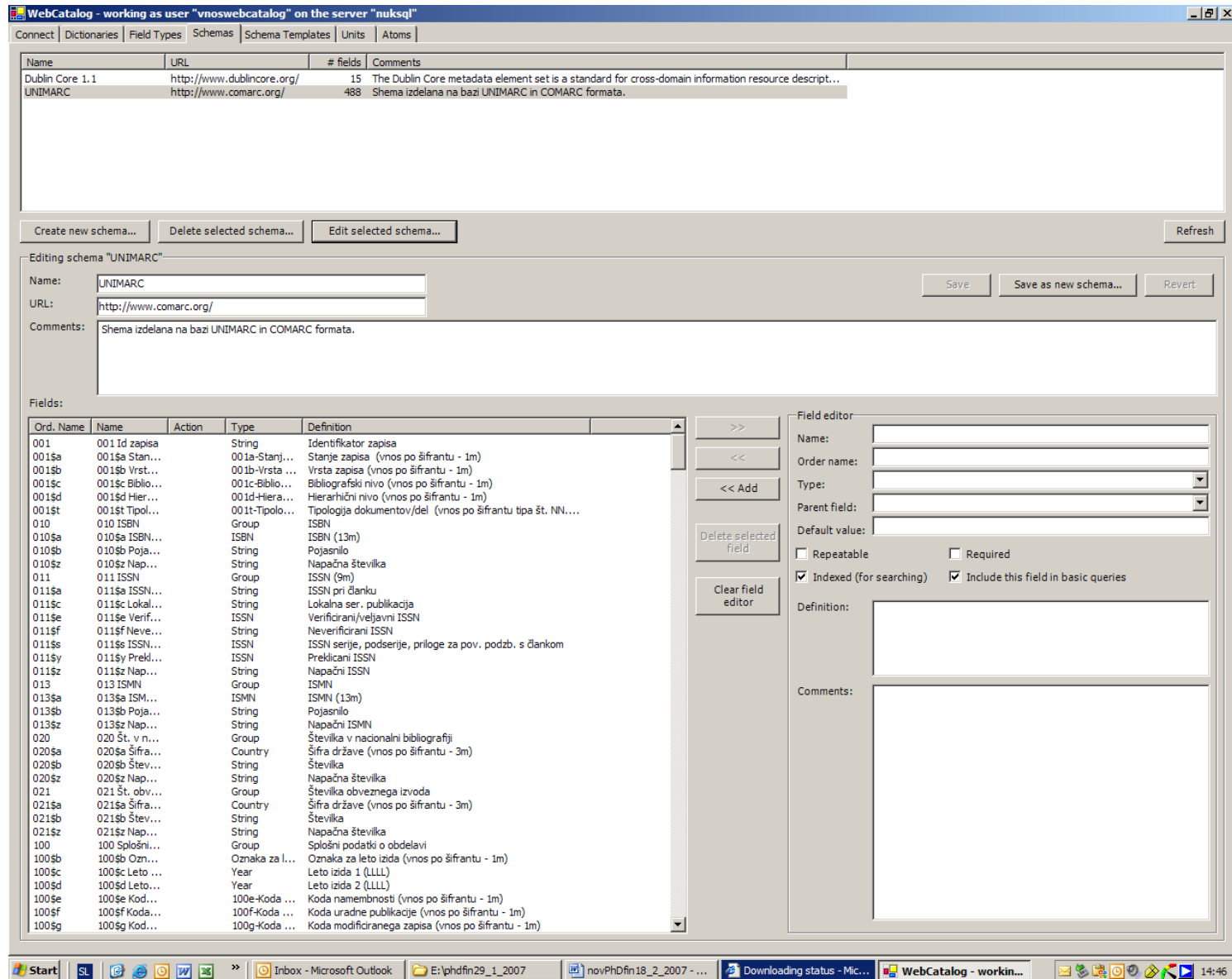
- Definirali smo sučelja lokalne bibliografske baze za unos i pregledavanje podataka o preuzetim mrežnim stranicama, koja su izradili istraživači IJS-a. Baza je vrlo fleksibilna i omogućava samostalno stvaranje sheme(a) za unos podataka. Njezina struktura s jedne strane prati logiku UNIMARC-a, a s druge rješava problem granularnosti mrežne građe. U njoj se može definirati tip polja ili potpolja te se mogu kreirati jednostavne maske za unos, kao što je Dublin Core, odnosno složenije maske u MARC formatima. Aplikacija podupire i unos popisa kodova za standardan opis jedinica građe. Razlika između jedinica građe i njezinih dijelova omogućava grupiranje po različitim kriterijima i stvaranje zajedničkoga bibliografskoga opisa. Sustav također omogućava unos identifikatora u opis građe. (vidi **slike br. 32-39**).



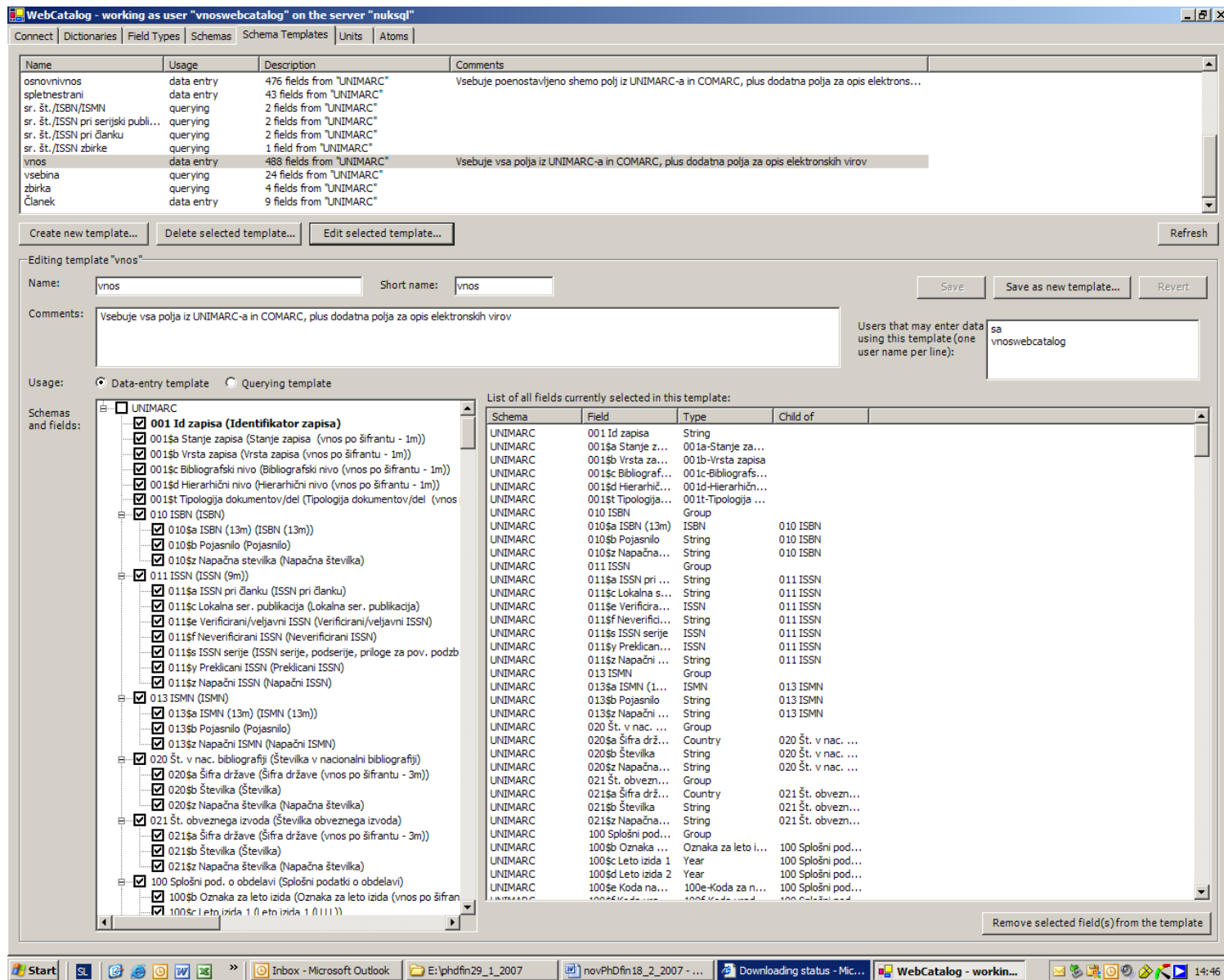
Slika br.32: Sučelje za definiciju koda u poljima



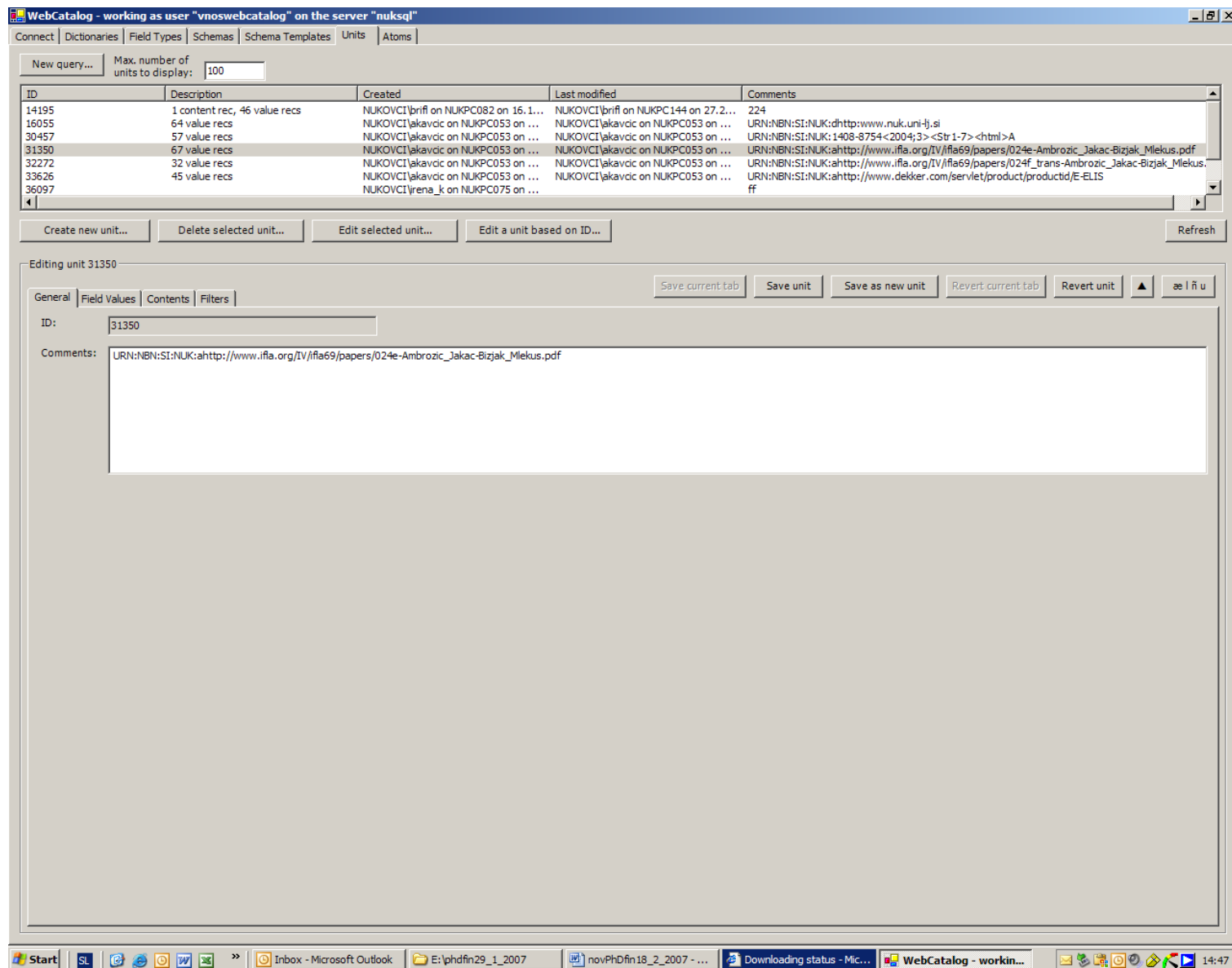
Slika br. 33: Sučelje za definiciju tipova polja



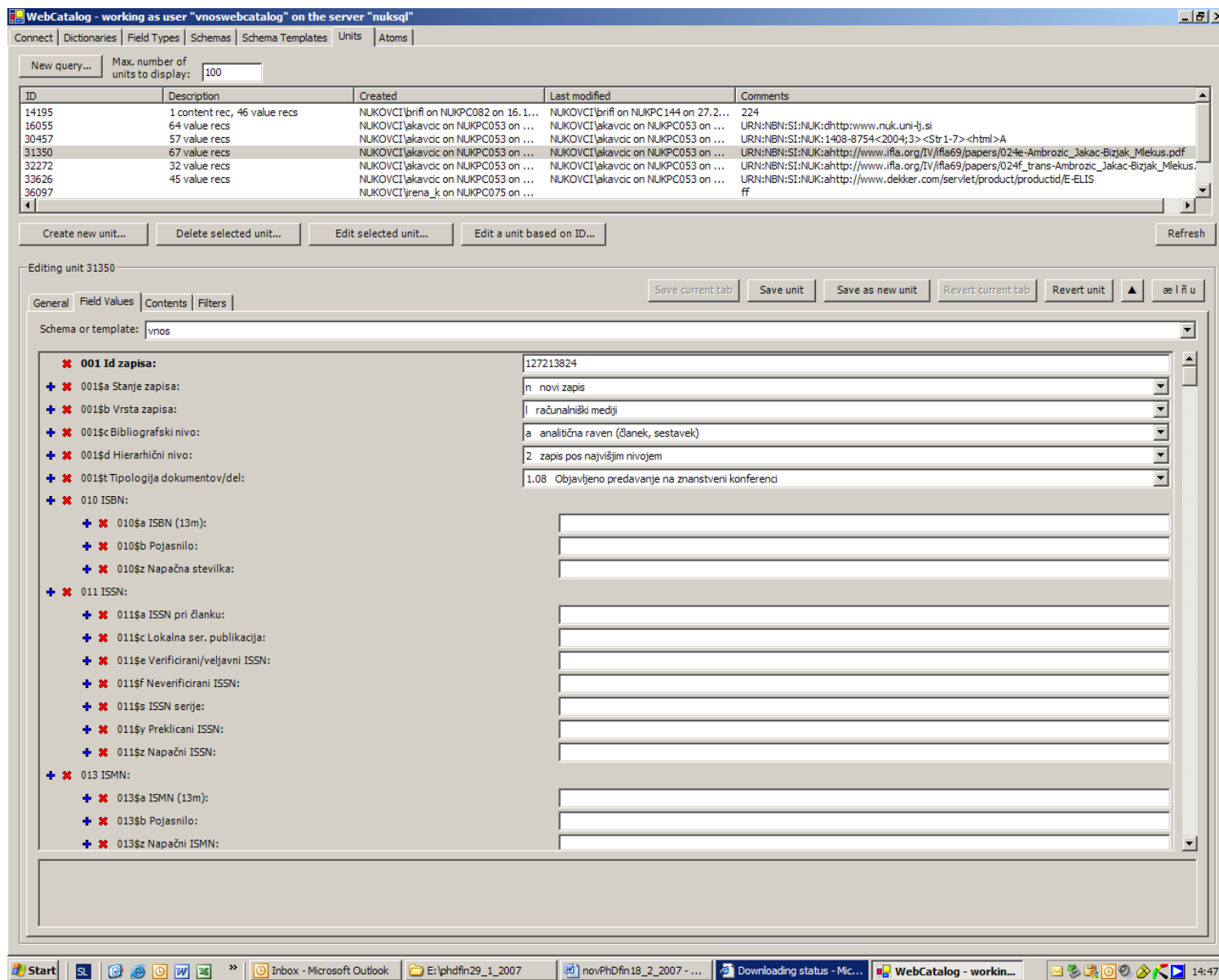
Slika br. 34: Sučelje za definiciju strukture bibliografskih formata



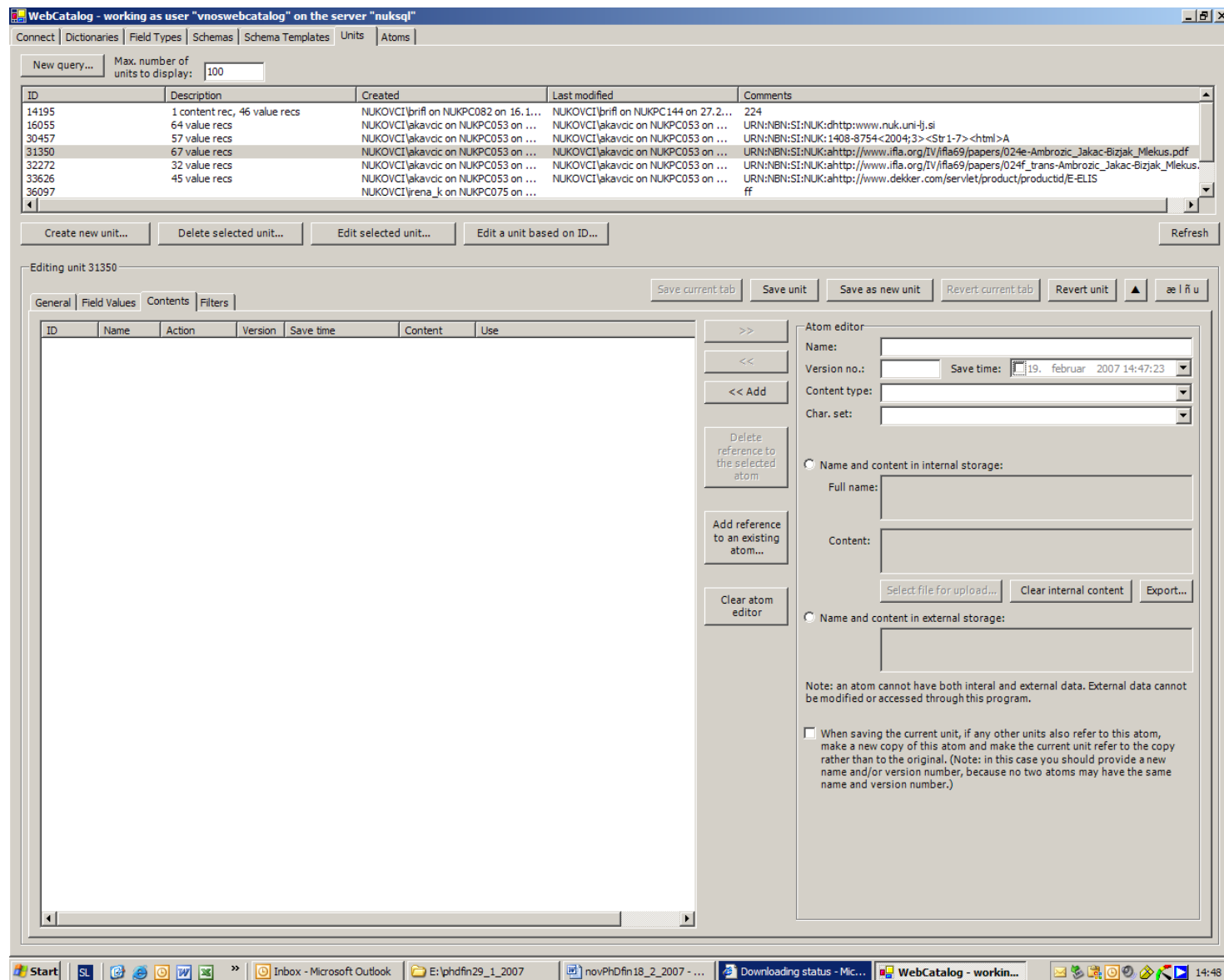
Slika br. 35: Sučelje za oblikovanje maske za unos podataka



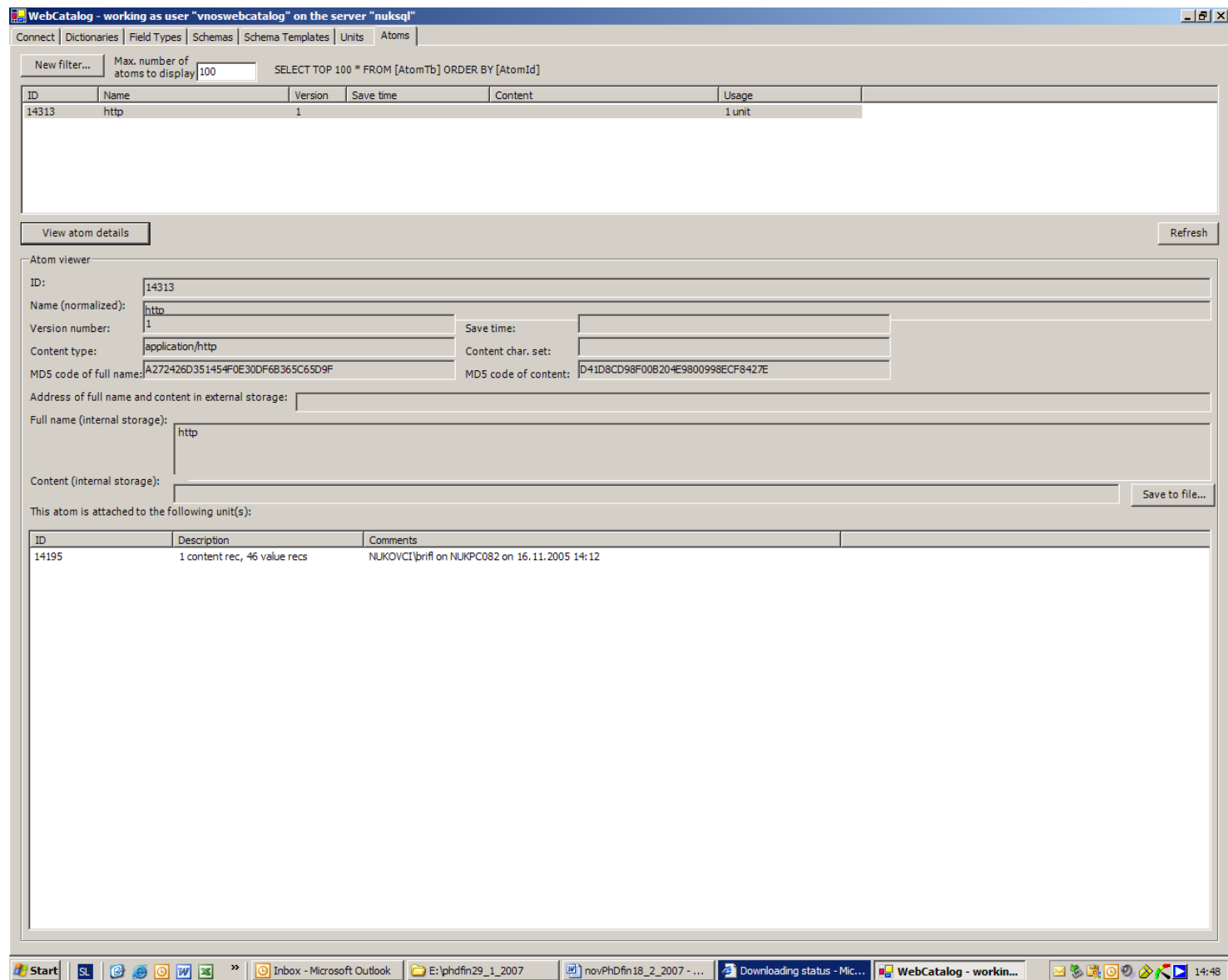
Slika br.36: Sučelje za opis digitalnih objekata - unos trajnih identifikatora



Slika br.37 : Sučelje za opis digitalnih objektov - vnosi bibliografskega opisa



Slika br.38: Sučelje za opis digitalnih objekata - maska za popravljanje zapisa



Slika br. 39: Sučelje za grupiranje mrežne građe u sadržajne jedinice s jednim bibliografskim opisom

Na posebnom radnom prostoru na poslužitelju pohranjuju se nove elektroničke publikacije koje bibliotekar pojedinačno preuzima, grupira i obrađuje te ih smješta u digitalni arhiv. Usto ovaj sustav omogućava da se iz web arhiva odvoje relevantni URL-i i obrade u skupinama ili pojedinačno. Uz to sučelje bilo je definirano i internetsko sučelje za pretraživanje po web arhivu, i to na tri načina: po bibliografskim podacima koji su smješteni u opisanoj bibliografskoj bazi; po punom tekstu i po vremenskim odrednicama.

6. Definiran je algoritam za stvaranje identifikatora URN. (vidi **dodatak br. 2**)
7. Definirano je sučelje za automatsko slanje/primanje mrežne građe (**dodatak br. 3**)
8. Definirane su operativne potrebe za prikupljanje mrežne građe te potrebe za stručnjacima i informacijskom tehnologijom te su predviđeni okvirni troškovi.
9. Pripremljen je informacijski koncept portala u vezi sa sadržajima digitalnoga arhiva i prikupljanjem digitalne građe. Glavne točke navedene su u prethodnom poglavlju.
10. Pripremljena su polazišta za gradnju digitalnoga arhiva i za izvođenje prakse prikupljanja mrežne građe. U cjelini su uključena u dokument "*Strategija NUK v zvezi z gradnjo digitalne knjižnice Slovenije - dLIB.si*".

Dok smo se u NUK-u posvetili razvoju metodologije, u IJS-u su programirali robot. Tako je nastao *WebBird Crawler*.³⁷¹ Napisan je u programskom jeziku C++ i djeluje u operacijskom sustavu *Windows 2000/XP*. Sadržaj dokumenata smješta se u posebno priređenom formatu koji omogućava jednostavan rad s podacima i rekonstrukciju baze u slučaju teškoća (naprimjer ako se indeksi poruše). Metapodatci se pohranjuju u posebnoj bazi podataka koja djeluje na platformi *Microsoft SQL Server*. Obje su baze, baza sadržaja i baza indeksa, povezane. Osim baze, izgrađen je i modul za indeksiranje podatka, koji indeksira sadržaje dokumenata u bazi podataka i metapodatke u relacijskoj bazi. Modul za traženje omogućava pretraživanje po toj relacijskoj bazi. Budući da su autori istraživači iz odjela semantičkih analiza i *text mining* metoda, u IJS-u su predvidjeli i modul za vizualizaciju dokumenata.

³⁷¹ Grobelnik, M. ; Mladenich, D. Poročilo projekta "Arhiviranje slovenskega spleta". // Kavčič-Čolić, A. ... [et al.]: Metodologija zbiranja in arhiviranja slovenskih elektronskih publikacij na medmrežju / [nositeljica projekta Kodrič-Dačić, E. ; projektna skupina Kavčič-Čolić, A. (rukovodilac projekta) ... et al.]. - Ljubljana : Narodna in univerzitetna knjižnica, 2004. 105 str. (Ciljni raziskovalni programi)

IJS je izradio prvu verziju robota u srpnju 2003. godine. Njegovo smo djelovanje testirali 15.7.2003. Odabrali smo uzorak od devetnaest mrežnih mjesta ili 50.150 mrežnih stranica. Uzorak je bio odabran na osnovi vrste nakladnika (javne i privatne organizacije: javna uprava, kulturne ustanove, sveučilišta, društveno-političke organizacije, komercijalna i uslužna poduzeća, nakladnici itd), sadržaja (portali, serijske publikacije, slike, statične i dinamične stranice) i strukture (jednostavna struktura s manjim brojem mrežnih stranica i složene strukture s više razina ili s datotekama u *Flashu* i/ili u *Java Scriptu*)

Analiza prikupljene mrežne građe pokazala je sljedeće nepravilnosti:

- kod većeg broja URL-a bilo je teško naći naslovnicu;
- ugrađene manje baze podataka, kao što su ankete u mrežnim mjestima političkih organizacija ili baze s pretraživačima, nisu djelovale;
- složeniije strukture nisu bile u cjelini zahvaćene ili su nastajale kružne klopke;
- kod nekih naslovnica s više razina nedostajale su stranice koje nisu u istoj domeni, a relevantne su;
- nije bilo jasno zašto su kod nekih domaćih stranica preuzete sve stranice s identičnom domenom, a kod nekih ne;
- prekinute su poveznice koje su upućivale na snimke u MP3 datotekama - vjerojatno su bile isključene iz definicije robota;
- datoteke u *Flashu* nisu djelovale;
- u mnogim primjerima pojavio se problem prikazivanja dugmad izbornika;

U prosjeku 68% od svih analiziranih mrežnih stranica bilo je dostupno. Neke su stranice bile 100% dostupne.

Zbog tih pogrešaka i problema razlikovanja apsolutnih i relativnih adresa URL-a u IJS-u su morali i dalje nastaviti dogradnju i razvoj softvera. U vrijeme pisanja ovoga rada istraživači IJS-a dorađuju u robotu funkcije za leksičko prepoznavanje mrežnih stranica u slovenskom jeziku.

12.1 Ocjena rezultata projekta

Poznato je da "tko radi taj i griješi", što se dogodilo i kod ovog projekta. Činjenica je da kad bismo ponovno započeli taj isti projekt sa sadašnjim iskustvima, drugačije bismo planirali neke aspekte. U nastavku je dana ocjena rezultata i ponuđeni su odgovori na otvorenih pitanja.

U projektu smo se usredotočili samo na četiri vrste publikacija na mreži, a morali bismo se usredotočiti na samu mrežu, njezinu promjenljivost i vrste publikacija koje u njoj nalazimo. To bi bila osnova za stvaranje nove tipologije mrežne građe.

Kriteriji selekcije bili su vrlo elementarni pa nismo obuhvatili svu građu, već samo četiri osnovna tipa građe. Iskustva na europskom projektu (*Cultural Web Heritage*) pokazala su da je problematika mnogo složenija.

Metapodatci za dugoročnu pohranu temeljili su se na rezultatima projekta NEDLIB. Danas bismo vjerojatno upotrijebili pragmatičniji pristup vodeći računa o stupnjevima granularnosti mrežne građe, naprimjer implementacija rješenja novozelandske Nacionalne knjižnice u njemačkoj Nacionalnoj knjižnici (LMER).

Na početku smo imali poteškoća kod definicije odnosa između postojećega skupnoga kataloga COBIB i digitalnoga arhiva. Uprava NUK-a željela je izgraditi samostalni digitalni arhiv i zato smo odlučili preslikati pojednostavljenu strukturu iz COMARC-a u bazu na SQL, koja bi poslužila i za unos bibliografskih podataka i druge mrežne građe. To je bila pogrešna odluka jer bismo se trebali koristiti postojećim bibliografskim bazama. Sada nam novi protokoli za izmjenu podataka (primjerice OAI-PMH) olakšavaju prijenos iz COBIB-a u ostale baze. To bismo morali iskoristiti za prijenos samo osnovnih podataka u lokalnu bazu (po ISO ili Dublin Core), a zatim bismo dodali sve ostale informacije koje su važne za dugoročnu pohranu. Alternativno je rješenje proširiti format COMARC lokalne baze COBIB. O tome još razmišljamo.

U okviru projekta nismo predvidjeli prijenos građe u XML. To je već uspješno riješeno u Hrvatskoj, Njemačkoj, Nizozemskoj, Kongresnoj knjižnici SAD-a i ostalim nacionalnim knjižnicama.

Još uvijek je URN optimalno rješenje za nacionalne knjižnice. Međutim, u projektu nismo predvidjeli sustav za razrješavanje URN-a s URL-ima.

Puno smo vremena posvetili autorskom pravu jer tada još nismo imali Zakon o obveznom primjerku i nismo znali kada će i hoće li uopće njegov prijedlog biti prihvaćen.

Digitalni smo repozitorij pokušali graditi istovremeno s web arhivom. Bilo bi pametnije da smo prvo pripremili glavni "kontejner" za digitalnu građu, a tek onda bi tražili način kako da tu građu prenosimo iz mreže. Problem je naravno u tome što su ti sustavi veoma skupi (implementacija Fedore u nacionalnoj knjižnici Estonije stajala je 200.000 €) i zahtijevaju skupinu programera i sistemskih inženjera.

Ne zaboravimo da je tada razvoj rješenja za dugoročnu pohranu bio tek u nastanku i sva rješenja za koja smo se odlučili bila su u tom trenutku optimalna. Kad bismo danas ponovno započinjali sličan projekt, pokušali bismo u najvećoj mjeri iskoristiti postojeće softverske aplikacije (naprimjer robot *Heritrix*), a ne bismo išli u samostalan razvoj. Istina je da preuzimanje tuđih rješenja stvara neku vrstu ovisnosti. Međutim, naša suradnja s istraživačkim institutom pokazala je, da istraživači drukčije razumiju aplikativne projekte i u želji da se postigne savršenstvo produžuju razvoj WebBirda do beskrajnosti. Dvije godine nakon formalnoga završetka opisanoga projekta još uvijek nemamo alat kojim bismo prikupljali slovensku mrežnu građu. Možda je doista vrijeme da počnemo ispočetka.

Projekt je zato bio vrlo uspješan na drugim područjima. Prvi put smo u NUK-u posvetili veliku pozornost procesima dugoročne pohrane. Neke od smjernica uključili smo u NUK-ovu strategiju iz 2004. godine i u izgradnju portala dLib.si.³⁷² Analizirali smo probleme prikupljanja i obradbe elektroničke građe i sve prateće metapodatke. Izvještaj o projektu također je bio osnova za

³⁷² Krstulović, Z. Slovenska Digitalna knjižnica - dLIB.si. Ljubljana : NUK, 2004.

pripremu i odlučivanje o Zakonu o obveznom primjerku u Ministarstvu za kulturu. Konačno, bez projekta bismo teško saznali gdje smo pogriješili i kojim područjima moramo posvetiti veću pozornost.

Budući da na pogreškama učimo, projekt "Metodologija skupljanja i arhiviranja slovenskih elektroničkih publikacija na mreži" bio je za NUK bogato iskustvo.

13. ZAKLJUČAK

Razvoj informacijskoga društva stvara uvjete za sve veću i širu upotrebu interneta kao komunikacijskoga i publikacijskoga kanala. Internet je postao dio svakodnevnoga života građana razvijenoga svijeta i kao takav važan je dokaz njihove društvene, kulturne i znanstvene djelatnosti. Svaki pojedinac s pristupom internetu postaje potencijalni autor. Na njemu se reproduciraju sadržaji u tradicionalnim oblicima, ali i stvaraju novi oblici sadržaja, koji nisu nikad prije postojali. Kao posljedica toga, s jedne strane broj mrežne građe neprestano raste i mijenja se, a s druge strane građa na mreži ima sve kraći životni ciklus. Osim toga dobiva i nove konvencionalne oblike koji konkuriraju već postojećoj tradicionalnoj građi na materijalnim nositeljima.

Glavno pitanje na koje smo u ovoj disertaciji tražili odgovor je "Kako sačuvati za buduće generacije svu relevantnu produkciju na internetu?". Odgovor je vrlo složen i pritom treba sagledati različite bitne aspekte koje metodologija za arhiviranje mrežne građe mora uzeti u obzir. Na osnovi najnovijih dostignuća i prakse na području arhiviranja mrežne građe razmotrili smo i prikazali moguća rješenja. Iskustva i praksa sagledani su iz perspektive potreba i situacije slovenske Narodne i sveučilišne knjižnice. Činjenica je da svaki pristup ima svoje prednosti i nedostatke. Optimalan rezultat može se postići samo kombinacijom više pristupa.

U disertaciji smo posvetili pažnju sljedećim problemima:

- Specifične karakteristike mrežne građe znatno utječu na način njihova prikupljanja, pohrane i pristupa. Mrežnu građu moramo shvatiti kao višedimenzionalni pojam, s različitom logikom, strukturom i dodatnim funkcijama kojih nema u tradicionalnoj građi na materijalnim nositeljima.
- Pri arhiviranju mrežne građe vrlo veliku ulogu imaju alati i metode za njihovo prikupljanje, kao i infrastruktura u kojoj se pohranjuje. Robot za prikupljanje mrežne građe mora biti sposoban izvoditi različite funkcije automatskoga i selektivnoga, odnosno ručnoga prikupljanja građe. Najnoviji razvoj na tom području koji vodi konzorcij IIPC obećava velik

napredak u pronalaženju dobrih rješenja u budućnosti. Svako rješenje ovisi o situaciji u organizaciji i o ciljevima pohrane.

- Pri izgradnji mrežnih arhiva treba imati u vidu to da dio mrežne građe sadrži svojstva zbog kojih je postojeći alati za prikupljanje građe ne mogu prikupiti. Taj dio nazivamo dubinskom mrežom i u njoj je veliki dio građe koju vrijedi očuvati.
- Postoji velika razlika između web arhiva i ostalih digitalnih arhiva u načinu pohrane i pristupa toj građi. U međunarodnoj su praksi su poznati osnovni uvjeti izgradnje digitalnih arhiva. Te uvjete po svojoj funkciji i organizacijskim strateškim ciljevima ispunjavaju nacionalne i druge depozitarne knjižnice.
- Postoje različite tipologije mrežne građe koje se u većini slučajeva temelje na tradicionalnim konceptima. Te se tipologije temelje prije svega na sadržaju, formi ili funkcionalnosti jedinica građe. Genologija interneta već nekoliko godina pokušava definirati nove sadržajne oblike publiciranja, odnosno žanrove, kao što su naslovnice, blogovi, baze podataka, elektroničke enciklopedije itd., koje je već prihvatila mlada generacija.
- Što je vrijedno sačuvati za budućnost? Hoće li možda za dvadeset godina biti vrijedniji blog koji obrađuje ekološku problematiku u regiji ili naslov znanstvene elektroničke serijske publikacije iz područja društvenih znanosti? Ne postoji način na koji bismo to u ovom trenutku saznali. Zato je važno da pomoću postojećih alata i metoda pokušamo pronaći optimalan način za dugoročnu pohranu većega dijela pisane baštine na mreži. Međutim, budući da ćemo uvijek biti ograničeni sredstvima ili vremenom, posebnu pozornost treba posvetiti kriterijima odabira. Ti su kriteriji najvažnija polazišna točka kako za namještanje funkcija robota, tako i za ručni odabir građe.

Svaka depozitarna institucija prilagođava svoje kriterije odabira svojim strateškim ciljevima, mogućnostima i okruženju. NUK svoje kriterije temelji na kriterijima slovenike. Cilj je većine nacionalnih knjižnica prikupiti svu vrijednu građu iz perspektive sadašnjosti. Trenutačno su to djela s nekom vrstom recenzije uredničkoga odbora ili znanstvenika s autoritetom. Velik je dio tih mrežnih djela pristupačan na komercijalnoj osnovi (elektronički časopisi, disertacije,

istraživački radovi, elektroničke knjige i slično). Novi oblici mrežne građe tek dobivaju svoj identitet i moguće je da će u budućnosti odigrati vrlo veliku ulogu.

Specifični kriteriji odabira slovenike moraju obuhvatiti sadržajne, formalne i tehničke perspektive pri ručnom i automatskom prikupljanju mrežne građe.

- U slovenskoj praksi Zakon o obveznom primjerku daje ovlasti NUK-u za prikupljanje i dugoročnu pohranu mrežne građe. Unatoč tome, autorsko je pravo vrlo važan element kod omogućavanja pristupa toj građi. Interesi nakladnika i autora kod komercijalnih dijela moraju se poštovati i mora se naći optimalan kompromis za korisnike. Moguće rješenje je omogućavanje lokalnoga pristupa komercijalnoj građi samo na jednom računalu ili određenom broju njih u NUK-u. Kodeks prakse dobrovoljnoga depozita može poslužiti kao smjernica.
- Bibliografska obradba mrežne građe morala bi biti usmjerena prema nalaženju relevantne literature s pozicije sadašnjosti. To znači da bi pratila klasične oblike i postupno uključivala nove žanrove, koji dobivaju svoju formu. Zbog bibliografske kontrole bibliografska obradba mora biti što potpunija. Za razliku od ručno preuzete elektroničke građe, automatski prikupljena građa zahtijeva nove pristupe obradbi. Njezino nam prikupljanje omogućava zahvaćanje ogledala društvenoga, kulturnoga, političkoga, znanstvenoga i povijesnoga trenutka našega društva. To su milijuni mrežnih stranica koje je nemoguće pojedinačno obraditi, a moramo ih sačuvati. Robot mora omogućiti da se proces prikupljanja informacija o toj građi obavi automatski.
- Postavlja se pitanje je li dovoljno da bibliotekarstvo prilagodi postojeća pravila novim oblicima publiciranja ili bi bilo jednostavnije ponovno započeti sve iz druge perspektive i iskoristiti sve mogućnosti i logičke strukture koje nam pruža računalna tehnologija. Jesu li stari koncepti, koji publikaciju uzimaju kao diskretnu formu s naslovnicom i impresumom, upotrebljivi i za mrežnu građu različitih hibridnih oblika, koja dobiva novi identitet u elektroničkom okruženju? Kako sačuvati kontekst, koji je u elektroničkom okruženju neodvojiv od sadržaja?

- Proces dugoročne pohrane, za razliku od tradicionalne građe, započinje vrlo rano, u trenutku njezina prikupljanja i unosa u digitalni arhiv, a po mogućnosti još i ranije, u procesu njezina stvaranja. Strategije za dugoročnu pohranu vrijede za svu građu. Neki se procesi mogu automatizirati pri ulasku u arhivski sustav. Međutim, za sve je procese vrlo važno da se planiraju i da su svi postupci dobro definirani. Koliko smo bili uspješni, znat ćemo tek za petnaest godina, odnosno kroz jednu generaciju. Toliko vremena predviđaju stručnjaci da je potrebno da bi se stara tehnologija u potpunosti zamijenila novom, s kojom dolazi i novo znanje. Međutim, stručnjaci su na području dugoročne pohrane optimistični i vjeruju da ćemo je, budemo li sustavno pratili postupke i procese pohrane elektroničke građe, uspjeti ćemo riješiti u budućnosti. Pritom je vrlo važna međunarodna suradnja, prije svega na području bibliografske obradbe. Zbog specifičnih svojstava elektronička građa zahtijeva dodatne informacije o njezinu djelovanju i prikazivanju, koje kod tiskanih publikacija nisu potrebne. Zato su metapodaci za dugoročnu pohranu vrlo važan element njihova opisa.
- S elektroničkom se građom pojavila potreba za trajnim identifikatorima, koji nam pomažu da identificiramo i lokaliziramo građu u elektroničkom okruženju. Među svim vrstama trajnih identifikatora trenutačno su URN-i najbolje rješenje za nacionalne knjižnice.
- Standardi i pravilnici na području dugoročne pohrane vrlo su važni jer daju mogućnost prijenosa i razmjene znanja i nalaženje optimalnih rješenja na tom području. Među njima je i ISO standard 14721 (teoretski model OAIS) koji su depozitarne ustanove prihvatile kao osnovu za izgradnju digitalnih i web arhiva. Budući da je taj model vrlo općenit i fleksibilan, može se prilagoditi različitim situacijama i okruženjima. U ovom je radu upotrijebljen model Titie van der Werf, koji bio izrađen u okviru projekta NEDLIB, a s njim je izrađen i sustav dugoročne pohrane u NUK-u. U modelu su opisani procesi koje možemo razumjeti u okviru prethodno analiziranih područja.
- Vrijedno iskustvo na području arhiviranja mreže bio je projekt NUK-a na kojem smo postavili prva polazišta za stvaranje metodologije arhiviranja slovenskih publikacija na mreži. Očekivani rezultat projekta, na kojem su sudjelovali i istraživači IJS-a morao bi biti aktivan

web arhiv i uspostavljanje prakse arhiviranja slovenskih publikacija na mreži. Do kraja 2006. godine ti rezultati nisu ostvareni u praksi.

U disertaciji su potvrđene dvije teze:

- Prva je teza glasila da je moguće izraditi metodologiju i metodološki model koji će usmjeriti knjižničare pri uspostavljanju sustava za arhiviranje slovenike na mrežnim stranicama. Svako od područja kojima smo se posvetili u ovom radu čini jedan važan aspekt arhiviranja mrežne građe, kao i osnovu metodološkoga modela. Procese prikupljanja, pohrane, arhiviranja i omogućavanje korisničkoga pristupa treba sagledati kao dio cjeline, dio sustava. Bilo koja metodologija mora posvetiti pozornost svim opisanim aspektima i pokušati naći optimalno rješenje u okviru postojećih mogućnosti.
- Druga je teza da taj metodološki model mora biti usklađen s referentnim modelom OAIS (ISO 14721), koji je zacrtao osnovni referentni model pri izradi elektroničkih arhiva. Zbog njegove fleksibilnosti postoje danas različite implementacije toga modela. Konzorcij IIPC pokušava ga upotrijebiti u web arhivu. U vrijeme pisanja ovoga rada IIPC još nije izradio praktičnu primjenu modela web arhivu. Inače, disertacija obrađuje mogućnost njegove konkretizacije na primjeru NUK-a. S tim u vezi u ovom smo radu izradili dijagrame za svaki proces posebno, u okviru cjelovitoga sustava za arhiviranje elektroničke građe. Uz svaki su proces navedeni svi važniji postupci i mogućnosti, ovisno o pristup ili metodi arhiviranja.

Na kraju još ostaju sljedeća otvorena pitanja:

- Hoće li razvoj robota za prikupljanje mrežne građe prevladati nove tehnologije i softvere koji su izrađeni u Javi i ostalim objektno orijentiranim jezicima ili dinamičnim mrežnim stranicama koje trenutačne verzije ne mogu prepoznati?
- Je li moguće automatizirati i bibliografsku obradbu mrežne građe do te mjere da više neće biti potrebna ručna intervencija?
- Hoće li semantičke mreže i ostale text-mining metode naći svoju primjenu u knjižnicama?

- Postoji potreba stvaranja nove tipologije elektroničkih publikacija. Koliko će ti oblici utjecati na međunarodne bibliografske standarde ili preporuke institucija kao što je IFLA ICABS i ostale?
- I konačno, hoćemo li uspjeti pohraniti pristup elektroničke pisane baštine u našim arhivima za sto i više godina?

Informacijsko-tehnološki razvoj odvija se brzo i generacije koje se nalaze na prijelazu u informacijsko društvo mora se cijelo vrijeme iznova prilagođavati promjenama. Te se promjene mogu osjetiti u suvremenom shvaćanju pojma publikacije, u preispitivanju starih koncepata koji su označili dugu tradiciju bibliotekarstva te u koegzistenciji s informacijskom tehnologijom. Promjene su u svakom slučaju potrebne. Međutim, kao akteri u samom razvoju teško postavljamo distancu koja bi nam omogućila da iznova preispitamo i prihvatimo nova pravila djelovanja. Tehnologija ne smije biti cilj, već mora biti alat za postizanje cilja. Kad bismo toga bili svjesni, već bismo davno iskoristili mogućnosti koje nam pružaju istraživanja na području semantičkih mreža i funkcionalnosti pretraživača, a bibliotekarstvo bi se koncentriralo na sadržaj i funkciju, a ne na formu.

14. LITERATURA

- Abiteboul, S. [et al.]: A first experience in archiving the French web. // Research and advanced technology for digital libraries : 6th European Conference, ECDL 2002, Rome, Italy, September 16-18, 2002 : Proceedings / uredili Agosti, M. ; Thanos, C. Berlin ... etc. : Springer, 2002. Str. 1-15.
- Abrams, S. L. The role of format in digital preservation. // VINE: The Journal of Information and Knowledge Management Systems. 34, 2(2004), str. 49-55.
- Albertsen, K. The Paradigma web harvesting environment. // 3rd Workshop on Web Archives, Trondheim, Norway, August 21st, 2003 : Proceedings / uredili Masanes, J. ; Rauber, A. ; Cobena, G. URL: <http://bibnum.bnf.fr/ECDL/2003/proceedings.php?f=albertsen> (15.12.2006.), str. 49-62.
- Andersen, B. The DK-domain: in words and figures, [2006]. URL: http://netarkivet.dk/publikationer/DFrevy_english.pdf. (15.12.2006)
- Anglo-American Cataloguing Rules. Second Edition, 2002 Revision / prepared under the direction of the Joint Steering committee for Revision of AACR, a committee of the American Library Association [et al.]. Chicago : American Library Association ; Ottawa : Canadian Library Association ; London : Chartered Institute of Library and Information Professionals, 2004.
- Arms, W. Y. Collecting and Preserving the Web: The Minerva Prototype. // RLDigiNews. 5, 2(2001). URL: <http://www.rlg.org/preserv/diginews/diginews5-2.html#feature1>. (15.12.2006.)
- Arvidson, A. The collection of Swedish web pages at the Royal library - The web heritage of Sweden. // 68th IFLA Council and General Conference: Libraries for Life: Democracy, Diversity, Delivery, August 18th - 24th 2002, Glasgow, Scotland : [proceedings]. URL: <http://www.ifla.org/IV/ifla68/papers/111-163e.pdf>. (15.12.2006.)
- Arvidson, A. The Kulturarw3 project - the Swedish Royal web archive // The Electronic Library. 16, 2(1998), str. 105-108. URL: <http://kulturarw3.kb.se/html/projectdescription.html>. (15.4.2003.)
- Aschenbrenner, A. Long-term preservation of digital material - building an archive to preserve digital cultural heritage from the Internet. Master Thesis, Technical University Vienna, Dec. 2001. URL: <http://www.ifs.tuwien.ac.at/~aola/publications/thesis-ando/Contents.html>. (15.12.2006.)
- Ashenfelder, M. Web harvesting and streaming media. // IAWW'06 : Proceedings of the 6th International Web Archiving Workshop, September 21-22 2006, Alicante, Spain / Masanès, J. ; Rauber, A., str. 125-145. URL: <http://www.iwaw.net/06/>. (15.12.2006.)
- Ashley, K. The preservation of databases. // VINE : The Journal of Information and Knowledge Management Systems. 34, 2(2004), str. 66-70.
- Asirvatham, A. P. ; Ravi, K. K. Web page categorization based on document structure, 2001. URL: <http://gdit.iiit.net/~arul/paper.pdf>. (15.12.2006.)

- Askehave, I. ; Nielsen, A. E. What are the characteristics of digital genres? - Genre theory from a multi-modal perspective // Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS'05) - Track 4, 2005. URL: <http://csdl2.computer.org/persagen/DLAbsToc.jsp?resourcePath=/dl/proceedings/&toc=comp/proceedings/hicss/2005/2268/04/2268toc.xml>. (15.12.2006.)
- An audit checklist for the certification of trusted digital repositories : draft for public comment. Mountain View, CA : RLG-OCLC, Aug. 2005. URL: <http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf>. (15.12.2006.)
- Baeza-Yates, R. ; Castillo, C. Crawling the infinite web, 2004. URL: http://www.dcc.uchile.cl/~ccastill/papers/baeza_05_crawling_infinite_web.pdf. (15.12.2006.)
- Bailey, S. ; Thompson, D. IKWAC : Building the UK's first public web archive. // D-Lib Magazine. 12, 1(2006). URL: <http://dlib.org/dlib/january06/thompson/01thompson.html>. (15.12.2006.)
- Baly, N. ; Sauvin, F. Archiving streaming media on the web : proof of concept and first results. // IWWA'06 : Proceedings of the 6th International Web Archiving Workshop, September 21-22 2006, Alicante, Spain / Masanès, J. ; Rauber, A. URL: <http://www.iwaw.net/06/>. (15.12.2006.), str. 147-181.
- Baron, N. S. Who wants to be a discipline? // The Information Society. 21 (2005), str. 269-271.
- Baym, N. K. Introduction: Internet research as it isn't, is, could be, and should be. // The Information Society. 21 (2005), str. 229-232.
- Bergman, M. K. The 'Deep' Web : Surfacing Hidden Value. White Paper. BrightPlanet. Sept. 2001. The Journal of Electronic Publishing August, 2001 Volume 7, Issue 1 URL: <http://www.brightplanet.com/images/stories/pdf/deepwebwhitepaper.pdf>. (19.2.2007.)
- Berners-Lee, T. ; Masinter, L. ; McCahill, M. RFC 1630 - Universal Resource Identifiers in WWW (URI), 1994. URL: <http://tools.ietf.org/html/rfc1630>. (3.11.2006.)
- Berners-Lee, T., Masinter, L. and McCahill, M.: RFC 1738 - Uniform Resource Locators (URL), 1994. URL: <http://rfc.dotsrc.org/rfc/rfc1738.html>. (3.11.2006.)
- Bibliotheca Universalis – Collections – NDL. URL: http://www.kb.nl/gabriel/bibliotheca-universalis/en/ndl/Bibliotheca%20Universalis_ndl.htm. (14.10.2005.)
- Blau, J. Do burned CDs have a short life span?. PC World. 10.1.2006. URL: <http://msn.pcworld.com/article/id,124312-page,1/article.html>. (15.12.2006.)
- Blood, R. The Weblog handbook : practical advice on creating and maintaining your blog. Cambridge MA : Perseus Publishing, 2002.
- Borbinha, J. ; Freire, N. ; Neves, J. BN.: the architecture of a National Digital Library // Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries, 2004. 7-11 June 2004. DOI:10.1109/JCDL.2004.1336090. URL: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=1336090&isnumber=29473. (15.12.2006.) , str. 21- 22

- Brown, A. Digital Preservation Guidance Note 2: Selecting storage media for long-term preservation. Richmond : The National Archives, 2003. (URL: <http://www.nationalarchives.gov.uk/preservation/advice/digital.htm>. (15.12.2006.)
- Brown, A. Digital Preservation Guidance Note 3: Care, handling and storage of removable media. Richmond : The National Archives, 2003. URL: <http://www.nationalarchives.gov.uk/preservation/advice/digital.htm>. (15.12.2006.)
- Brown, A. Archiving websites : a practical guide for information management professionals. London: Facet Publishing, 2006.
- Brygfjeld, S. A. Access to web archives: the Nordic Web Archive Access Project. // 68th IFLA Council and General Conference: Libraries for Life: Democracy, Diversity, Delivery, August 18th - 24th 2002, Glasgow, Scotland: [proceedings]. URL: [URL:http://www.ifla.org/IV/ifla68/papers/090-163e.pdf](http://www.ifla.org/IV/ifla68/papers/090-163e.pdf). (17.4.2003.)
- Buckland, M. K. What is a "Document". // Journal of the American Society for Information Science. 48, 9(1997), str. 804-809.
- Cantara, L. METS : The Metadata Encoding and Transmission Standard. // Cataloging & Classification Quarterly. 40, 3-4(2005), str. 237-253.
- Cathro, W. ; Webb, C. ; Whiting, J. Archiving the web : the PANDORA Archive at the National Library of Australia. A paper presented by Warwick Cathro at the Preserving the present for the future web archiving conference, Copenhagen, 18-19 June 2001. URL: <http://www.nla.gov.au/nla/staffoaoer/2001/cathro3.html>. (15.12.2006.)
- Cedars Guide to preservation metadata, 2002. URL: <http://www.leeds.ac.uk/cedars/guideto/metadata>. (15.12.2006.)
- CENL: http://www.nlib.ee/cenl/working_groups.php. (15.12.2006.)
- Christensen, N. H. Towards format repositories for web archives. // 4th IAWW 2004. URL: <http://www.iwaw.net/04/Christensen.pdf>. (15.12.2006.)
- Christensen-Dalsgaard, B. ... [et al.]. Experiences and conclusions from a pilot study: web archiving of the district and country elections 2001. Final report for the pilot project "netarkivet.dk", February 2003. <http://netarchive.dk/publikationer/webark-final-rapport-2003.pdf>. (15.12.2006.)
- Christensen-Dalsgaard, B. Web archive activities in Denmark. // RLG DigiNews. 8, 3(2004). URL: http://www.rlg.org/en/page.php?Page_ID=17661#article0. (15.12.2006.)
- Code of practice for the voluntary deposit of electronic publications. // Conference of European National Librarians/Federation of European Publishers (CENL/FEP), 2000. URL: <http://www.fep-fee.be/cenlen.htm> (15.12.2006.)
- Collection Policy / National Library of New Zealand: URL: <http://www.natlib.govt.nz/en/about/1keypolcollections.html>. (15.12.2006)
- Collections policy statement : web site capture & archiving. Washington: Library of Congress, 2004. URL: <http://www.loc.gov/acq/devpol/webarchive.html>. (15.12.2006.)
- COMBINE. Lund: NetLab, 1999. URL: <http://www.lub.lu.se/combine/>. (15.4.2003.).

- Crowston, K. ; Williams, M. Reproduced and emergent genres of communication on the world wide web. // The Information Society. 16 (2000), str. 207-208.
- Data Dictionary for Preservation Metadata : Final Report of the PREMIS Working Group. Dublin (OH) : OCLC ; Mountain View (CA): RLG, 2005.
- Day, M. CEDARS: Digital preservation and metadata. // Sixth DELOS Workshop: Preservation of Digital Information, Tomar, Portugal, 17-19 June 1998. URL: <http://www.ukoln.ac.uk/metadata/presentations/delos6/cedars.html>. (5.4.2004.)
- Delsey, T. Preispitivanje konvencionalnih paradigmi za opis dokumenata = Reassessing conventional paradigms for document description / prevela Mirna Willer // Vjesnik bibliotekara Hrvatske. 46, 1-2(2003), str. 32-43.
- Digital Curation Centre (DCC). URL: <http://www.dcc.ac.uk/resource/technology-watch/dspace>. (15.12.2006.)
- The Digital Dilemma: Intellectual Property in the Information Age. // Committee on Intellectual Property Rights and the Emerging Information Infrastructure, the Computer Science & Telecommunications Board, and the Commission on Physical Sciences, Mathematics, and Applications. Washington : National Academy Press, 2000. URL: http://bob.nap.edu/html/digital_dilemma/. (15.12.2006.)
- Digital Preservation Coalition Handbook: Chapter 5: Media and Formats. URL: <http://www.dpconline.org/graphics/medfor/media.html>. (15.12.2006.)
- DINI-Certificate. Document and Publication Repositories. Göttingen : Deutsche Initiative für Netzwerkinformation, Oct. 2003. URL: <http://www.dini.de/documents/Zertifikat-en.pdf>. (7.2.2004.)
- Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. // Official Journal. L 077, 27/03/1996 str. 0020-0028. URL: <http://europa.eu.int/eur-lex/lex/Notice.do?val=344298:cs&lang=en&list=391654:cs,391398:cs,276720:cs,276918:cs,284595:cs,266982:cs,272215:cs,236779:cs,344298:cs,415100:cs,&pos=9&page=2&nbl=27&pgs=10&hwords=directive~databases~&checktexte=checkbox&visu=#texte>. (15.12.2006.)
- Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society. // Official Journal. L 167, 22/06/2001, str. 0010-0019, URL: http://eur-lex.europa.eu/LexUriServ/site/en/oj/2001/l_167/l_16720010622en00100019.pdf. (15.12.2006.)
- ECUP: Position on user rights in electronic publications. // The Librarian Quarterly. 7(1997), str. 386-395.
- Elektronske publikacije. Kodeks prakse prostovoljnega depozita / [tekst priredila i dopunila Vilenka Jakac-Bizjak]. Ljubljana : Narodna in univerzitetna knjižnica, 2001 (Ljubljana : Pleško).
- Emigh, W. ; Herring, S. C. Collaborative authoring on the web: a genre analysis of online encyclopedias. // Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS'05) - Track 4, 2005.

- <http://csdl2.computer.org/comp/proceedings/hicss/2005/2268/04/22680099a.pdf>. (15.12.2006.).
- Etični kodeks slovenskih knjižničarjev // ZBDS, 1995. URL: <http://www.nuk.uni-lj.si/nuk3.asp?id=37119387>
 - FAIFE. URL: <http://www.faiife.dk/> (15.12.2006.)
 - Falk, H. Digital archive developments. The Electronic Library. 21, 4(2003), str. 375-379.
 - 14th Statistical Survey Report on the Internet Development in China. China Internet Network Information Center (CNNIC), 20 July 2004. URL: <http://www.cnnic.net.cn/html/Dir/2004/07/22/2406.html>. (15.12.2006)
 - Functional Requirements for Authority Records: A conceptual model. Draft 2005-6-15. // IFLA UBCIM Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). URL: <http://www.ifla.org/VII/d4/FRANAR-Conceptual-M-Draft-e.pdf>. (15.12.2006.)
 - Functional Requierments for Bibliographic Records : Final Report. // IFLA Study Group on the Functional Requirements for Bibliographic Records. München : K.G. Saur, 1998. (UBCIM Publications - New Series Vol. 19). - URL: <http://www.ifla.org/VII/s13/frbr/frbr.pdf>. (15.12.2006.)
 - Gambari, S., Guerrini, M. Definire a catalogare le risorse elettroniche. Milano : Editriche Bibliografica, 2002.
 - Garret, J. Waters, D. Preserving Digital Information. Report of the Task Force on Archiving of Digital Information (1996). Commission on Preservation and Access and RLG. URL: <http://www.rlg.org/legacy/ftpd/pub/archtf/final-report.pdf>. (7.2.2004)
 - Gatenby, P. Digital archiving - developing policy and best practice guidelines at the National Library of Australia. URL: <http://www.icsti.org/icsti/2000workshop/gatenby.html>. (15.4.2003.)
 - Garvas Delić, A. ; Milinović, M. Prikupljanje, odabir, obrada i pohrana dokumenata sa službenih mrežnih stranica tijela javne vlasti Republike Hrvatske. // 2. Kongres hrvatskih arhivista, Dubrovnik, listopad 2005. godine / Hrvatsko arhivističko društvo ; Državni arhiv u Dubrovniku. URL: http://www.hidra.hr/hidra/10godina/POOPD_za_HAD.pdf. (19.2.2007.)
 - Gomes, D. ; Silva, M. J. A characterization of the Portuguese web. // 3rd Workshop on Web Archives, Trondheim, Norway, August 21st, 2003 : Proceedings / uredili Masanès, J. ; Rauber, A. ; Cobena, Gregory, 2003. Str. 63-76. URL: <http://bibnum.bnf.fr/ecdl/2003/index.html>. (15.12.2006.)
 - Gomes, D. ; Silva, M. On URL and content persistence. Prosinac 2005, str. 5. URL: <http://www.di.fc.ul.pt/tech-reports/05-21.pdf>. (15.12.2006.)
 - Grobelnik, M. ; Mladenič, D. Poročilo projekta "Arhiviranje slovenskega spleta". // Kavčič-Čolić, A. ... [e tal.]: Metodologija zbiranja in arhiviranja slovenskih elektronskih publikacij na medmrežju / [nositeljica projekta Kodrič-Dačić, E. ; projektna skupina Kavčič-Čolić, A. (rukovodilac projekta) ... et al.]. - Ljubljana : Narodna in univerzitetna knjižnica, 2004. (Ciljni raziskovalni programi).

- Gross, J. Learning by doing : the Digital Archive for Chienes Studies (DACHS). // 3rd Workshop on Web Archives, Trondheim, Norway, August 21st, 2003 : Proceedings / uredili Masanès, J. ; Rauber, A. Cobena, G. URL: <http://bibnum.bnf.fr/ecdl/2003/proceedings.php?f=gross>. (15.12.2006), str. 103-110.
- Guidelines for the preservation of digital heritage // prepared by the National Library of Australia. Paris : UNESCO, 2003. <http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>. (21.9.2004)
- Guilliland-Swetland, A. J. Introduction to Metadata: Setting the Stage, 2000. (Pathways to Digital Information). URL: <http://www.getty.edu/research/institute/standards/intrometadata>. (5. 5. 2004.)
- Gulli, A. ; Signorini, A. The indexable web is more than 11.5 billion pages. // ACM. URL: <http://www.cs.uiowa.edu/~asignori/pubs/web-size/>. (15.12.2006.)
- Habert, B. ; Hurault-Plantet, M. ; Jardino, M. Web page classification: field categories and functional categories. SENSNET RNRT Project, 2005. URL: <http://www.limsi.fr/RS2005/chm/lir/lir4/>. (15.12.2006.)
- Hakala, J. Archiving the web : European experiences. // Program: electronic library and information systems, 38, 3(2004), str. 176-183.
- Hakala, J. Archiving the Web: European experiences. // Tietolinja. 2(2003). URL: <http://www.lib.helsinki.fi/tietolinja/0203/webarchive.html> (15.12.2006.)
- Hakala, J. Collecting and preserving the web : developing and testing the NEDLIB harvester. // RLG DigiNews. 5, 2(2001). URL: <http://www.thames.rlg.org/preserv/diginews/diginews5-2.html> (15.4.2003), str. 2.
- Hakala, J. Internet metadata and library cataloguing. // International Cataloguing & Bibliographic Control. 28. 1(1999), str.21-25.
- Hallgrímsson, Þ. ; Bang, S. Nordic web archive. // IAWAW 2003. URL: <http://bibnum.bnf.fr/ecdl/2003/proceedings.php?f=hallgrimsson>. (15.12.2006.)
- Harper, G. Will we need fair use in the twenty-first century? 2001. URL: http://www.utsystem.edu/OGC/Intellectualproperty/fair_use.htm. (15.12.2006.)
- Hendry, D. G. ; Carlyle, A. Hotlist or bibliography? A case of genre on the web. // Proceedings of the 39th Hawaii International Conference on System Sciences (HICSS'06) - Track 3, 2006. <http://csdl2.computer.org/comp/proceedings/hicss/2006/2507/03/250730051b.pdf>. (15.12.2006.)
- Heritrix. URL. <http://crawler.archive.org/>. (15.12.2006.)
- Herring, S. C. ... [et al.]. Bridging the gap : a genre analysis of weblogs // Proceedings of the 37th Hawaii International Conference on System Sciences (HICSS'04) - Track 4, 2004. URL: <http://www.ics.uci.edu/~jpd/classes/ics234cw04/herring.pdf> (19.2.2007.)
- Hilse, H.W. ; Kothe, J. Implementing persistent identifiers : overview of concepts, guidelines and recommendations. London : Consortium of European Research Libraries (CERL) ; Amsterdam : European Commission on Preservation and Access (ECPA), 2006.

- Hine, C. Internet research and the sociology of cyber-social-scientific knowledge. // The Information Society. 21 (2005), str. 239-248.
- Horvat, T. E-arhiv službenih publikacija i dokumenata tijela javne vlasti RH. // 35. skupština Hrvatskoga knjižničarskog društva, Plitvička jezera, Hotel Jezero 27.-30.9.2006. URL: <http://www.hidra.hr/hidra/10godina/plitvice.pdf>. (19.2.2007.)
- How much information 2003. URL: <http://www2.sims.berkeley.edu/research/projects/how-much-info/> (15.12.2006.).
- Information and Documentation : The WARC File Format (working draft proposed to ISO TC46/SC4). URL: <http://www.niso.org/international/sc4/N595.pdf>. (15.12.2006.)
- ISBD(CR) : International standard bibliographic description for serials and other continuing resources : revised from ISBD(S) : international standard bibliographic description for serials / recommended by the ISBD(S) Working Group ; approved by the Standing Committees of the IFLA section on Cataloguing and the IFLA Section on Serial Publications. München : Saur, 2002. (UBCIM publications. New series, ISSN 0941-8946 ; Vol. 24).
- ISBD(CR) : međunarodni standardni bibliografski opis serijskih publikacija i druge neomeđene građe : prerađeno izdanje ISBD(S): Međunarodnoga standardnoga bibliografskog opisa serijskih publikacija / [priređivač] Međunarodni savez knjižničarskih društava i ustanova ; [s engleskog prevela Ana Barbarić]. Zagreb : Hrvatsko knjižničarsko društvo, 2005. (Povremena izdanja Hrvatskog knjižničarskog društva, ISSN 1334-7667. Novi niz ; knj.11).
- ISBD(ER) : međunarodni standardni bibliografski opis elektroničke građe : prerađeno izdanje ISBD(CF)-a: Međunarodnoga standardnoga bibliografskog opisa računalnih datoteka / preporučila Radna grupa za pregled ISBD(CF)-a ; [s engleskog prevele, hrvatske primjere odabrale i izradile Tanja Buzina i Sofija Klarin]. Zagreb : Hrvatsko knjižničarsko društvo, 2001. (Povremena izdanja Hrvatskog knjižničarskog društva. Novi niz ; knj. 3)
- HTTrack. URL: <http://www.httrack.com>. (15.12.2006.)
- Izjava o međunarodnih katalogizacijskih načelih: osnutek, sprejet na 1. srećanju strokovnjakov za izradu međunarodnoga katalogizacijskoga pravilnika u okviru IFLE, Frankfurt, Njemačka, 2003 / prevele Kavčič, I. i Kalčić, D. URL: http://www.ddb.de/standardisierung/pdf/statement_slovene.pdf. (15.12.2006.)
- Jacobsen, G. Universal Bibliographic Control and International MARC Core Programme (UBCIM) : International Conference on National Bibliographic Services Copenhagen, 25-27 November 1977: COVERAGE. // ICNBS, Copenhagen, 25-27 November 1998. URL: <http://www.ifla.org/VI/3/icnbs/fina.htm>. (15.12.2006.)
- Jakac-Bizjak, V. Problematika avtorskih pravic v zvezi z zbiranjem in hranjenjem elektronskih publikacij ter z njihovim dostopnostjo za javnost. // Kavčič-Čolić, A. ... [e tal.]. Metodologija zbiranja in arhiviranja slovenskih elektronskih publikacij na medmrežju / [nositeljica projekta Eva Kodrič-Dačić ; projektna skupina Alenka Kavčič-Čolić (rukovodilac projekta) ... et al.]. Ljubljana : Narodna in univerzitetna knjižnica, 2004. (Ciljni raziskovalni programi).
- Jansen, H. Permanent access : the e-Depot at the Koninklijke Bibliotheek. URL: <http://www.nla.gov.au/webarchiving/JansenHans.rtf>. (15.12.2006.)

- Jørgensen, P. H. Practical application of FRBR and RDF. // *Le risorse elettroniche : definizione, selezione e catalogazione = Electronic resources. Definition, selection and cataloguing. Proceedings of the International Conference, Rome, November 26-28, 2001 / uredio Guerrini, M. Milano : Bibliografica, 2002. str. 623.*
- Kavčič-Čolić, A. Archiving the web - some legal aspects. // 68th IFLA Council and General Conference, August 18-24, 2002. Također objavljeno u: *Library Review*. 52, 5(2003), str. 203-208.
- Kavčič-Čolić, A. Metapodatki za trajno ohranjanje elektronskih virov. // *Knjižnica*. 48, 4(2004), str. 97-119.
- Kavčič-Čolić, A. Teoretični model digitalnih arhivov. // *Knjižnica*. 48, 4(2004), str. 63-75.
- Kavčič-Čolić, A. ; Grobelnik, M. Archiving the Slovenian web : recent experiences. // *Proceedings / 4th International Web Archiving Workshop (IWA04), held in conjunction with the 8th European conference on research and advanced technologies for digital libraries, September 16 2004, Bath, UK. Bath : University, 2004. URL: <http://www.iwaw.net/04/index.html>. (15.12.2006.)*
- Kavčič-Čolić, A. ... [et al.]. Evaluation report : Deliverable D No. 3.2. eContent Programme reUSE - digital master files of printed publications : eContent Project No. 11173. Ljubljana : NUK, 2006. URL: <http://www.uibk.ac.at/reuse/documents/>. (21.9.2004)
- Kavčič-Čolić, A. ... [et al.]. Metodologija zbiranja in arhiviranja slovenskih elektronskih publikacij na medmrežju / [nositeljica projekta Kodrič-Dačić, E. ; projektna skupina Kavčič-Čolić, A. (rukovodilac projekta) ... et al.]. - Ljubljana : Narodna in univerzitetna knjižnica, 2004. (Ciljni raziskovalni programi).
- Kimpton, M. ; Ubois J. Year-by-Year: From an Archive on the Internet to an Archive on the Internet. // *Web archiving / uredio Masanès, J. Berlin : Heidelberg ; Springer, 2006. Str. 201-212*
- Klarin, S. ; Murati, T. Identifikacija, odabir, obradba i osiguravanje dostupnosti mrežnih publikacija u kontekstu nacionalne bibliografske kontrole // 4. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u kontekstu globalne informacijske infrastrukture. Zagreb : Hrvatsko knjižničarsko društvo, 2001. Str. 41-56.
- Klarin, S. ; Pigac, S. ; Pavelić D. Croatian remote access electronic serials: results of a survey. // *International Cataloguing & Bibliographic Control*, 30, 4(2001), str. 70-72.
- Kodrič, N. Analiza slovenskih elektronskih serijskih publikacij in njihovih založnikov. Priloga 2: // Kavčič-Čolić, A. ... [et al.]: Metodologija zbiranja in arhiviranja slovenskih elektronskih publikacij na medmrežju / [nositeljica projekta Kodrič-Dačić, E. ; projektna skupina Kavčič-Čolić, A. (rukovodilac projekta) ... et al.]. - Ljubljana : Narodna in univerzitetna knjižnica, 2004. 105 str. (Ciljni raziskovalni programi).
- Kodrič-Dačić, E. Slovenika: strokovne novosti novega Zakona o knjižničarstvu // *Knjižnica*, 46, 4(2002), str. 65-85
- Koerbin, P. The PANDORA Digital Archiving System (PANDAS): managing web archiving in Australia: a case study. Paper presented at the 4th International Web Archiving Workshop on 16 September 2004 in Bath, UK.

- Koerbin, P. Report on the crawl and harvest of the whole Australian web domain undertaken during June and July 2005, Okt. 2005. URL: http://pandora.nla.gov.au/documents/domain_harvest_report_public.pdf. (15.12.2006.)
- KOPAL. URL: <https://kopal.langzeitarchivierung.de/index.php.en>. (15.12.2006.)
- Krimbacher, M. ; Neuhauser, M. ; Vogl, M. Survey on the long-term preservation of digital documents in European Libraries. Innsbruck: Krimbacher Neuhauser & Vogl, 2005. http://www.uibk.ac.at/reuse/docs/reuse_survey_cd-rom.zip. (15.12.2006.)
- Krishnamurthy, S. The Multidimensionality of blog conversations: the virtual enactment of September 11. Maastricht, The Netherlands: Internet Research 3.0, 2002.
- Krstulović, Z. Digitalna knjižnica : izziv in prihodnost knjižnic = Digital library : the challenge and future of libraries // Informacijski viri in storitve knjižnic v elektronskem okolju / Strokovno posvetovanje Zveze bibliotekarskih društev Slovenije, Portorož, 24.-26. oktober 2005 = Professional Conference of Union of Associations of Slovene Librarians, Portorož, October 24-26, 2005. - Ljubljana : Zveza bibliotekarskih društev Slovenije, 2005. Str. 19-36.
- KULTURARW3. Stockholm : Royal Library (KB), Sweden's National Library. URL: <http://www.kb.se/kw3/ENG/Default.htm>. (15.12.2006.)
- Kuny, T. The digital dark ages? Challenges in the preservation of electronic information // International preservation news. 17, May (1998). URL: <http://www.ifla.org/VI/4/news/17-98.htm#2> (15.12.2006.)
- Kunze, J. A. Towards electronic persistence using ARK identifiers, 2002. URL: <http://www.cdlib.org/inside/diglib/ark/arkcdl.pdf>. (3.11.2006)
- Kunze, J. A. WARC: an archiving format for the web. // IWAW 2005. URL: <http://www.iwaw.net/05/kunze.pdf>. (22.9.2006)
- Kurahashi, T. Collection development of the National Diet Library. // CDNLAO Newsletter. 53(2005). URL: <http://www.ndl.go.jp/en/publication/cdnlao/053/534.html>. (14.10.2005.)
- Lamos, C., [et al.] : Archiving the Greek web. // 4th International Web Archiving Workshop on 16 September 2004 in Bath, UK, 2004. URL: <http://www.iwaw.net/04/Lamos.pdf>. (15.12.2006.)
- Lavoie, B. Meeting the challenges of digital preservation: The OAIS Reference Model // OCLC Newsletter. 243(2000) str. 26-30. <http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000001747&reqid=110130>
- Lecher, H. E. Small scale academic web archiving : DACHS. // Web archiving / uredio Masanès, J. Berlin Heidelberg : Springer, 2006. Str. 215.
- Lian'en, H. ; Hongfei, Y. ; Xiaoming, L. Engineering of Web InfoMall: The Chinese Web Archive // World Engineers' Convention 2004, November 2-6, 2004, Shanghai, China. URL: <http://net.pku.edu.cn/~yhf/refpaper/hle2004/l.jpg - 6.jpg>. (15.12.2006)
- Library of the University of Chicago. URL: <http://www.lib.uchicago.edu/e/net/types.html>. (15.12.2006.)

- Long-term preservation metadata for electronic resources (LMER). / uredio Steinke, T. Frankfurt am Main : Die Deutsche Bibliothek, 2005. URL: <http://www.ddb.de/eng/standards/lmer/lmer.htm>. (15.2.2007.)
- Lorie, R. A. A Project on Preservation of Digital Dana. // RLG DigiNews. 5, 3(2001). URL: <http://www.rlg.org/preserv/diginews/diginews5-3.html#feature2>. (15.12.2006.)
- Lupovici, C. Web crawling : The Bibliotheque nationale de France experience. // World Library and Information Congress: 71th IFLA General Conference and Council "Libraries - A voyage of discovery", August 14th-18th 2005, Oslo, Norway. URL: <http://www.ifla.org/IV/ifla71/papers/074e-Lupovici.pdf>. (15.12.2006.)
- Lupovici, C. ; Masanès, J. Metadata for the long-term preservation of electronic publications. The Hague : NEDLIB Consortium, 2000. (NEDLIB Report Series; 2), URL: <http://nedlib.kb.nl/results/NEDLIBmetadata.pdf>. (15.12.2006.)
- Lyman, P. ... [et al.]: How much information. URL: <http://www.sims.berkeley.edu/how-much-info/>. (15.12.2006.)
- MADS. URL: <http://www.loc.gov/standards/mads/>. (15.12.2006.)
- Mannerheim, J. The WWW and our digital heritage - the new preservation tasks of the library community // 66th IFLA Council and General Conference, Jerusalem, Israel, 13-18 August 2000. URL: <http://www.ifla.org/IV/ifla66/papers/158-157e.htm>. (15.12.2006.)
- Martins, B. ; Silva, M. Language identification in web pages. // Symposium on Applied Computing: Proceedings of the 2005 ACM symposium on Applied computing, 13.-17. ožujka 2005, Santa Fe, New Mexico. New York : ACM. -URL: http://xldb.di.fc.ul.pt/data/Publications_attach/ngram-article.pdf. (15.12.2006.)
- Masanès, J. IIPC Web Archiving Toolset. Netpreserve.org // IWAW 2005. URL: <http://www.nla.gov.au/webarchiving/MasanésJulien.ppt>. (15.12.2006.)
- Masanès, J. Web archiving // Digital preservation / uredili Deegan, M. ; Tanner, S. London : Facet Publishing, 2006. Str. 78-97.
- Masanès, J. Web archiving methods and approaches: a comparative study. // Library trends, 54, 1(2005), str. 72-90.
- Masanès, J. ; Stack, M. WERA : a web archive collection access tool. PPT presentation at IWAW 2005, Vienna, 22 Sept. 2005. URL: <http://www.iwaw.net/06/PDF/iwaw06-proceedings.pdf>. (15.12.2006.)
- Merriam-Webster Online Dictionary. URL: <http://www.m-w.com/cgi-bin/dictionary?book=Dictionary&va=encyclopedia&x=18&y=9>. (9.09.2005.)
- Metadata for digital preservation: the CEDARS project outline specification. Draft for public consultation / The Cedars Project Team and UKOLN, 2000. URL: <http://www.leeds.ac.uk/cedars/MD-STR~5.pdf>. (15.12.2006.)
- Metadata standards framework - preservation metadata (2002). Wellington: National Library of New Zealand, 2002. URL: http://www.natlib.govt.nz/files/4initiatives_metaschema.pdf. (15.12.2006.)

- METADISS. Frankfurt : Die Deutsche Bibliothek. URL: <http://deposit.ddb.de/metadiss.htm> (17.4.2003.)
- METS. URL: <http://www.loc.gov/standards/mets/>. (15.12.2006.)
- Milinović, M. ; Topolščak, N. The architectre of DAMP : a system for harvesting and archivnig web publications. // Widwisawn. 3, 3(2005). URL: http://widwisawn.cdrl.strath.ac.uk/Issues/Vol3/issue3_3_1.html. (15.12.2006.)
- MINERVA (Mapping the Internet Electronic Resources Virtual Archive). Washington : Library of Congress. URL: <http://www.loc.gov/minerva/>. (15.4.2003.).
- Mladenić, D. ; Grobelnik, M. ; Kavčič-Čolić, A. Initiatives to preserve Slovenian digital heritage // Innovation and knowledge economy: issues, applications, case studies / uredili Paul Cunningham and Miriam Cunningham. - Amsterdam [etc.] : IOS Press, 2005. (Information and communication technologies and the knowledge economy). Str. 993-998.
- MODS. URL: <http://www.loc.gov/standards/mods/>. (15.12.2006.)
- Mohr, G. ...[et al.]. An introduction to Heritrix : an open source archival quality web crawler. // 4th International Web Archiving Workshop (IWAW 2004). URL: <http://www.iwaw.net/04/Mohr.pdf>. (15.12.2006.)
- Monberg, J. Science and technolgoy studies approaches to internet research. // The Information Society. 21 (2005), str. 281-284.
- Müller, E. ... [et al.] Archiving Workflow between a local repository and the national archive. Experiences for the DiVa Project. URL: http://epc.uu.se/files/archiving_ECDL_2003.pdf. (15.12.2006)
- National Library to Capture New Zealand's Digital Heritage. URL: <http://www.natlib.govt.nz/bin/media/pr?item=1085885702>. (15.12.2006);
- National Library to lead electronic harvesting. URL: <http://www.natlib.govt.nz/bin/media/pr?item=1064531843>. (15.12.2006);
- National Library Of Canada, October 1998. URL: <http://www.collectionscanada.ca/9/8/index-e.html>. (15.12.2006)
- National Library of New Zealand - Digital Library Development Review. URL: http://www.natlib.govt.nz/files/ross_report.pdf. (15.12.2006)
- National Library of New Zealand (Te Puna Matauranga o Aotearoa) Act 2003. URL: <http://www.natlib.govt.nz/files/Act03-19.pdf>. (15.12.2006);
- NEDLIB (Networked European Deposit Library) [informacije o projektu]. Pridobljeno 5.4.2004 s spletne strani <http://nedlib.kb.nl/>. (5.4.2004.)
- NEDLIB Harvester. Den Haag : Koninklijke Bibliotheek, [s.i.]. URL: <http://www.csc.fi/sovellus/nedlib/> (15.4.2006).
- Nordic Web Archive. Helsinki : NORDINFO. URL: <http://nwa.nb.no/>. (15.4.2003.).
- Nordic Web Archive. Introduction. Appendix II - Specification for an archiving robot. Str. 15-18. URL: <http://www.lib.helsinki.fi/tietolinja/0100/nwa.pdf>. (15.12.2006.)

- NORDINFO (Nordic Council for Scientific Information). URL: <http://www.nordinfo.helsinki.fi/nordinfo/index.htm>. (15.12.2006.).
- NORDUnet2: URL: <http://www.nordunet2.org> (15.12.2006.)
- Nuys, C. van ; Albertsen, K. Identification of network accessible documents: problem areas and suggested solutions // 3rd ECDL Workshop on Web Archives, August 21st, 2003 Trondheim, Norway. URL: <http://bibnum.bnf.fr/ecdl/2003/index.html>. (15.12.2006.)
- O'Neill, E. T.: Characteristics of web accessible information. Presented at 63rd IFLA General Conference, August 31-September 5, 1997. // IFLA Journal. 24, 2(1998) <http://wotan.liu.edu/doi/data/Articles/julksrnqay:1998:v:24:i:2:p:7224.html>. (15.12.2006.)
- O'Neill, E. T. ; Lavoie, B. F. Bibliographic control for the web. // The Serials Librarian. 37, 3(2000), str. 53-69.
- Online Australian Publications: Selection Guidelines for Archiving and Preservation by the National Library of Australia. Revised August 2005. URL: <http://pandora.nla.gov.au/selectionguidelines1003.html>. (15.12.2006.)
- Orlikowski, W.J. ; Yates, J. Genre repertoire: the structuring of communicative practices in organizations. // Administrative Sciences Quarterly. 33(1994), str. 541-574.
- PADI - Web archiving: URL: <http://www.nla.gov.au/padi/topics/92.html#NZ>. (15.12.2006.);
- Pandora Business Process Model, URL: <http://pandora.nla.gov.au/bmp.html>. (15.12.2006.) (prvi put objavljen 1997. godine, obnovljen 2005. godine).
- Pedley, P. Copyright for library and information service. London : Aslib, 1998.
- Pedley, P. The invisible web : searching the hidden parts of the internet. London : ASLIB-IMI, 2001.
- Pênkavová, P. E-serials processing methodology. // Newsletter of the IFLA Section on Serial Publications. 39(2001), str. 12.
- Pennock, M. DSpace digital repository software, 12.6.2006. URL: <http://www.dcc.ac.uk/resource/technology-watch/dspace>. (15.12.2006.)
- Pennock, M. Fedora digital repository software, 24.11.2006. URL: <http://www.dcc.ac.uk/resource/technology-watch/fedora/>. (15.12.2006.)
- Phillips, M. E. Selective archiving of web resources: a study of acquisition costs at the National Library of Australia. // RLG DigiNews. June 15, 2005, URL: http://www.rlg.org/en/page.php?Page_ID=20666&Printable=1&Article_ID=1749. (15.12.2006.)
- Pollach, I. Electronic word of mouth: a genre analysis of product reviews on consumer opinion web sites. // Proceedings of the 39th Hawaii International Conference on System Sciences (HICSS'06) - Track 3, 2006. URL: <http://csdl2.computer.org/comp/proceedings/hicss/2006/2507/03/250730051c.pdf>. (15.12.2006.)
- Popović-Bošković, G. ; Fortuna, B. ISBD(CR) : Izzivi katalogizacije virov v nadaljevanju. // Organizacija znanja. 8, 2(2003).

- Popović-Bošković, G. ; Fortuna, B. Novi mednarodni katalogizacijski pravilnik : Korporativna značnica med idejo o enotni in univerzalno kontrolirani obliki // Organizacija znanja. 9, 2(2004). URL: http://home.izum.si/COBISS/OZ/2004_2/html/clanek_02.html. (15.12.2006.)
- PREMIS. URL: <http://www.oclc.org/research/projects/pmwg/>. (7.2.2005.)
- Preservation Metadata and the OAIS Information Model. A metadata framework to support the preservation of digital objects / OCLC/RLG Working Group on Preservation Metadata, 2002. URL: http://www.oclc.org/research/projects/pmwg/pm_framework.pdf. (5. 5. 2004.)
- Preservation metadata for digital collections : exposure draft. Canberra: National Library of Australia, 1999. URL: <http://www.nla.gov.au/preserve/pmeta.html>. (5. 5. 2004)
- Prime-Claverie, C. ; Beigbeder, M. ; Lafouge, T. Transposition of the cocitation method with a view to classifying web pages. // Journal of the American Society for Information Science and Technology. 55, 14(2004), str.1282-1289.
- PRONOM. URL: <http://www.nationalarchives.gov.uk/PRONOM/Software/proSoftwareSearch.asp?status=new> i <http://www.nationalarchives.gov.uk/pronom/about.htm>. (15.12.2006.)
- Prototype Evaluation: Web Archiving Project (WARP). Appendix 15. URL: <http://www.crl.edu/PDF/PCWAappendices1-20.pdf>. (14.10.2005.)
- Rauber, A. ; Aschenbrenner, A. ; Witvowt, O. Austrian online archive processing : analyzing archives of the world wide web. // Research and advanced technology for digital libraries : 6th European Conference, ECDL 2002, Rome, Italy, September 16-31, 2002 : Proceedings / uredili Agosti, M ; Thanos, C. Berlin ... etc. : Springer, 2002. Str. 1-15.
- Reference Model of an Open Archival Information System (OAIS) : Recommendation for space data system standards : CCSDS 650.0-B-1.Blue Book. // Committee on Intellectual Property Rights and the Emerging Information Infrastructure, the Computer Science & Telecommunications Board, and the Commission on Physical Sciences, Mathematics, and Applications. Washington: CCSDS (Consultative Committee for Space Data Systems), 2002. URL: http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html. (5.4.2004.)
- Reitz, J. M. Dictionary for library and information science. Westport : Libraries Unlimited, 2004. Str. 216.
- Resource Description Framework (RDF). URL: <http://www.w3.org/RDF/>. (15.12.2006)
- reUSE: <http://reuse.uibk.ac.at/>. (21.9.2004)
- Revising AACR2 to ACcommodate Seriality, 1999: Report to the Joint Steering Committee for Revision of AACR, April 1999. URL: <http://www.collectionscanada.ca/jsc/ser-rep4.html>. (3.11.2006.)
- Reynolds, R. ISSN, identifikatori i metapodaci u digitalnom svijetu = ISSN, identifiers and metadata in the digital workd. // Vjesnik bibliotekara Hrvatske. 47, 1-2(2004), str. 16-24.
- RFC 2141: URN Syntax: <http://www.faqs.org/rfcs/rfc2141.html> <<http://www.ietf.org/rfc/rfc2141.txt>>. (15.12.2006)

- RFC 2288 - Using Existing Bibliographic Identifiers as Uniform Resource Names: <<http://www.faqs.org/rfcs/rfc2288.html>>. (15.12.2006)
- RFC 3044 - Using The ISSN (International Serial Standard Number) as URN (Uniform Resource Names) within an ISSN-URN Namespace: <<http://www.faqs.org/rfcs/rfc3044.html>>. (15.12.2006)
- RFC 3061 - A URN Namespace of Object Identifiers: <<http://www.faqs.org/rfcs/rfc3061.html>> (15.12.2006)
- RFC 3187 - Using International Standard Book Numbers as Uniform Resource Names: <<http://www.faqs.org/rfcs/rfc3187.html>>. (15.12.2006)
- RFC 3188 - Using National Bibliography Numbers as Uniform Resource Names: <<http://www.faqs.org/rfcs/rfc3188.html>>. (15.12.2006)
- Ridi, R. The world as will and documentation. Definition, selection and access to remote electronic resources (RER). // Le risorse elettroniche : definizione, selezione e catalogazione = Electronic resources. Definition, selection and cataloguing. Proceedings of the International Conference, Rome, November 26-28, 2001 / uredio Guerrini, M. Milano : Bibliografica, 2002. Str. 439-451.
- Le risorse elettroniche : definizione, selezione e catalogazione : atti del convegno internazionale, Roma, 26-28 novembre 2001 / a cura di Mauro Guerrini con la collaborazione di Stefano Gambari e Lucia Sardo. Milano : Editrice Bibliografica, 2002. (Bibliografia e biblioteconomia)
- Samuelson, P. and Davis, R. The digital dilemma: a perspective on intellectual property in the information age. // 28th Annual Telecommunications Policy Research Conference, 2000. URL: <http://www.sims.berkeley.edu/~pam/papers/digdilsyn.pdf>. (15.12.2006.)
- Schneider, S. M. ... [et al.]. Building thematic web collections : challenges and experiences from the September 11 web archive and the Election 2002 web archive. // 3rd Workshop on Web Archives, Trondheim, Norway, August 21st, 2003 : Proceedings / uredili Masanès, J. ; Rauber, A. ; Cobena, G. URL: <http://bibnum.bnf.fr/ecdl/2003/proceedings.php?f=schneider>. (15.12.2006.), str. 77-93.
- Scott-Wilson, E. Identifiers and interoperability // Information architecture: designing information environments for purpose / uredili Glichrist, A. ; Mahon, B. London : Facet, 2004. (Managing information for the knowledge economy series / Series Editor: Angela Abell). Deseto poglavlje, str. 161-173.
- Seadle, M. METS and the metadata marketplace. // Library Hi Tech. 20, 3(2002), str. 255-257.
- Shepherd, M. ; Watters, C. The functionality attribute of cybergenres. // Proceedings of the 32nd Hawaii International Conference on System Sciences (HICSS'99), 1999. URL: <http://csdl2.computer.org/comp/proceedings/hicss/1999/0001/02/00012007.PDF>. (15.12.2006.)
- Shepherd, M. ; Watters, C. Identifying web genre : hitting a moving target. <http://users.cs.dal.ca/~watters/www2004WorkShop/pdfs/4.pdf>. (15.12.2006.)

- Shrum, W. Internet indiscipline: two approaches to making a field. // The Information Society. 21 (2005), str. 273-275
- Sigurd'sson, K. Incremental crawling with Heritrix. // 5th International Web Archiving Workshop (IWA05), 22-23 September 2005, Vienna, Austria. URL: <http://www.iwaw.net/05/papers/iwaw05-sigurdsson.pdf>. (15.12.2006.)
- Slats, J. ; Bakker, H. XML and digital preservation: Digital preservation testbed White Paper. Den Haag : Digital Preservation Testbed Project, 2002. URL: http://www.digitaleduurzaamheid.nl/bibliotheek/docs/white-paper_xml-en.pdf. (15.12.2006.)
- Smith, C. Building an Internet Archive System for the British Broadcasting Corporation. // Library Trends. 54, 1(2005), str. 16-32.
- Stack, M. Full text search of web archive collections. // 5th International Web Archiving Workshop (IWA05), 22-23 September 2005, Vienna, Austria. URL: <http://archive-access.sourceforge.net/projects/nutch/iwaw/iwaw-wacsearch.pdf>. (19.2.2007.)
- Stack, M. Heritrix crawler update. PPT presentation presented at 5th International Web Archiving Workshop (IWA05), 22-23 September 2005, Vienna, Austria. URL: <http://www.iwaw.net/05/stack1.pdf>. (15.12.2006.)
- Stančić H. Očuvanje elektroničkih informacijskih objekata : arhivi, knjižnice, muzeji - zajednička koncepcija. // 7. Seminar arhivi, knjižnice, muzeji : mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova / uredila Tinka Katić. Zagreb : Hrvatsko knjižničarsko društvo, 2004. (Izdanja Hrvatskoga knjižničarskog društva ; knj. 34). str. 26-35.
- Steenbakkens, J. The Nedlib Guidelines : Setting up a Deposit System for Electronic Publications. Den Haag: Koninklijke Bibliotheek, 2000. (NEDLIB Report series ; 5). URL: <http://www.kb.nl/coop/nedlib/results/NEDLIBguidelines.pdf>. (5.4.2004.)
- Steenbakkens, J. Setting up a Deposit System for Electronic Publications: The NEDLIB Guidelines. The Hague : Koninklijke Bibliotheek, 2000.
- Steinke, T. The universal object format - an archiving and exchange format for digital objects. // ECDL 2006 / Gonzalo, J. ... [et al.] (Eds.) , 2006. (LNCS 4172). Str 552-554.
- Stokes, J. H. Understanding Moore's Law, 2003. URL: <http://arstechnica.com/articles/paedia/cpu/moore.ars/1>. (15.12.2006.)
- Strong, W. S.: Copyright in the new world of electronic publishing. // presented at the workshop Electronic Publishing Issues II at the Association of American University Presses (AAUP) Annual Meeting, June 17, 1994, Washington, D.C. URL: <http://www.press.umich.edu/jep/works/strong.copyright.html>
- Summary of the questionnaire "Application of persistent identifiers (PI) at national libraries", Die Deutsche Bibliothek, 26.1.2004. URN: <http://www.urn-info.org/?link=300>. (15.12.2006)
- Šolar, R. Digitalno kartografsko gradivo, nov izziv kartografskih zbirk // Knjižnica. 47, 4(2003), str. 7-22.
- Thelwall, M.; Vaughan, L. A fair history of the Web? Examining country balance in the Internet Archive. // Library & Information Science Research. 26(2004), str. 162-176.

- Thibodeau, K. L. Overview of technological approaches to digital preservation and challenges in coming years. URL: <http://www.clir.org/pubs/reports/pub107/thibodeau.html>. (15.12.2006.)
- Trusted Digital Repositories: Attributes and responsibilities : an RLG-OCLC Report. Mountain View, CA : RLG, 2002. URL: <http://www.rlg.org/longterm/repositories.pdf>. (21.9.2004.)
- Turner, L. Digging deeper : the deep web // Book Report. 19, 5(2001), str. 43-46.
- Universal declaration of human rights. URL: <http://www.lonweb.org/onu/hr-ing-slo.htm>
- Verhoeven, I. H. Archiving web publications. Amsterdam : IBM Netherlands; The Hague : Koninklijke Bibliotheek, 2002. (IBM/KB Long-Term Preservation Study Report Series ; 6). URL: http://www.kb.nl/hrd/dd/dd_onderzoek/reports/6-webpublications.pdf. (15.12.2006.)
- Wagner, R. G. ; Gross, J. Harvesting the web, preserving Chinese voices. The Digital Archive for Chinese Studies (DACHS), 11 Nov. 2004. URL: <http://www.sino.uni-heidelberg.de/dachs/publications/taipei2004.rtf>. (15.12.2006)
- Wall, Raymond A.: Copyright made easier. 2nd ed. London : Aslib, 1998. - str. 338
- Web Archiving Project (WARP), URL: <http://warp.ndl.go.jp/>. (15.12.2006)
- WebArchiv: arhiv českého webu. <http://www.webarchiv.cz/> (15.12.2006)
- Web Characterization Project, URL: <http://www.oclc.org/research/projects/archive/wcp/default.htm> (15.12.2006)
- Web Cultural Heritage. URL: <http://www.webarchiv.cz/culture-2000-documents/>. (15.12.2006.)
- Web Cultural Heritage : Culture 2000 project 25.9.2005-24.9.2006. Prague : National Library of the Czech Republic, 2006. (15.12.2006.)
- Web sites : concepts, issues, and definitions. Dublin, OH : OCLC, January 1999. URL: <http://www.oclc.org/research/projects/archive/wcp/pubs/rn1-websites.htm>. (15.12.2006.)
- Werf, T. van der. The Deposit System for Electronic Publications: A process Model. Den Haag: Koninklijke Bibliotheek, 2000. (NEDLIB Report Series ; 6). URL: <http://www.kb.nl/coop/nedlib/results/DSEPprocessmodel.pdf>. (5.4.2004.)
- Weston, P. G. Between bibliographic and intellectual control. Some reflections on electronic bibliographic systems. // Le risorse elettroniche : definizione, selezione e catalogazione = Electronic resources. Definition, selection and cataloguing. Proceedings of the International Conference, Rome, November 26-28, 2001 / izdao Guerrini, M. Milano : Bibliografica, 2002. - str. 609-621.
- Wget. URL: <ftp://ftp.gnu.org/gnu/wget/>. (15.12.2006.)
- Wiggins, B. IFLA survey on inclusion of electronic resources in national bibliographies. // World Library and Information Congress: 71th IFLA General Conference and Council "Libraries - A voyage of discovery", August 14th-18th 2005, Oslo, Norway.
- Willer, M. ; Milinović, M. DAMP: sustav za preuzimanje i arhiviranje obveznog primjerka hrvatskih mrežnih publikacija. // Tehnički in vsebinski problemi klasičnega in elektronskega

- arhiviranja : zbornik referatov dopolnilnega izobraževanja s področij arhivistike, dokumentalistike in informatike. 4(2005). Str. 491-500.
- Winer, D. The history of weblogs, 2002. URL: <http://oldweblogscomblog.scripting.xom/historyOfWeblogs>. (15.12.2006.)
 - Yan, H. ... [et al.]: A new data storage and service model of China web InfoMall. // 4th International Web Archiving Workshop, 2004. URL: www.iwaw.net/04/proceedings/Hongfei.pdf. (15.12.2006)
 - Yates, J. ; Orlikowski, W. J. Genres of organizational communication: a structural approach to studying communication and media. // The Academy of Management Review. 17, 2(1992), str. 299-326.
 - Zakon o avtorski in sorodnih pravicah - uradno prečiščeno besedilo // Uradni list RS. 94(2004). URL: http://zakonodaja.gov.si/rpsi/r08/predpis_ZAKO4148.html (15.12.2006.)
 - Zakona o knjižnicama // Narodne novine. 105(1997); 5(1998); 104(2000)
 - Zakon o knjižničarstvu // Uradni list Socialistične Republike Slovenije. 39, 27(1982), str. 1747-1752.
 - Zakon o knjižničarstvu // Uradni list Republike Slovenije, 11, 87(2001), str. 8685-8693.
 - Zakon o obveznem izvodu publikacij // Uradni list RS.69(2006)
 - Žabička, P. Archiving the Czech web: issues and challenges. // 3rd Workshop on Web Archives, Trondheim, Norway, August 21st, 2003 : Proceedings / uredili Masanès, J. ; Rauber, A. ; Cobena, G. URL: <http://bibnum.bnf.fr/ecdl/2003/proceedings.php?f=zabicka> (15.12.2006), str. 111-117.
 - Živković, D. Hrvatski ured za ISBN i ISMN : izveštaj o radu 1.1.-14.11.2002.

Popis tabela

	Str.
Tabela br. 1: Statistički podatci o broju mrežnih mjesta 1998.-2002. (Izvor: WCP, OCLC)	12
Tabela br. 2: Postotak preživjelih mrežnih mjesta po godinama	12
Tabela br. 3: Životni ciklus različitih materijalnih nositelja	149
Tabela br. 4: Shema metapodataka za dugoročnu pohranu elektroničke građe po CEDARS-u	174
Tabela br. 5: Izbor metapodataka za dugoročnu pohranu u projektu NEDLIB	178
Tabela br. 6: Izbor metapodataka OCLC i RLG	180
Tabela br. 7: Metapodatci za dugoročnu pohranu elektroničke građe u australskoj Nacionalnoj knjižnici	182
Tabela br. 8: Dodatna polja u COMARC-u za metapodatke za dugoročnu pohranu	225

Popis slika

	Str.
Slika br. 1: Dubinska mreža	15
Slika br. 2: Sustav koji razvija IIPC za web arhiv	64
Slika br. 3: Arhitektura Heritrixa	66
Slika br. 4: Anatomija datoteke WARC	70
Slika br. 5: Odnos cijene i kvalitete pri automatskom i ručnom prikupljanju mrežne građe	74
Slika br. 6: Primjer ekstenzivnoga prikupljanja mrežne građe, kod kojeg nisu bile uključene mrežne stranice a3 i c6	75
Slika br. 7: Primjer intenzivnoga preuzimanja mrežne građe	76
Slika br. 8: Implementacija različitih pristupa preuzimanja mrežne građe u funkciji učestalosti mijenjanja i interaktivnosti mrežnih stranica	78
Slika br. 9: Odnos građe slovenike, nacionalne bibliografije i obveznoga primjerka	111
Slika br. 10: Struktura elektroničke građe	147
Slika br. 11: Dvije različite logičke razine jednog dokumenta	147
Slika br. 12: Scenarij 1. arhiviranja podataka	152
Slika br. 13: Scenarij 2 arhiviranja podataka	153
Slika br. 14: Struktura informacijskoga paketa	158
Slika br. 15: Informacijski model OAIS	160
Slika br. 16: Funkcionalni entiteti OAIS	163
Slika br. 17: Depozitarni sustav za elektroničke publikacije	166
Slika br. 18: Primjer modela slojeva dugoročne pohrane elektroničke građe	168

Slika br. 19: Osnovni model slojeva za dugoročnu pohranu	169
Slika br. 20: Informacije o načinu prikazivanja elektroničke građe (representation information)	176
Slika br. 21 : Izbor metapodataka za dugoročnu pohranu elektroničke građe u Novom Zelandu	184
Slika br. 22: URN:NBN upotrijebljen je kao dio konvencije imenovanja datoteka i direktorija u arhivu DiVA	192
Slika br. 23: Koncept arhivskoga sustava u NUK-u	195
Slika br. 24: Prikaz procesa u arhivskom sustavu	197
Slika br. 25: Proces A1: Preuzimanje i prihvrat mrežne građe	202
Slika br. 26: Proces A2: Unos elektroničke građe u arhivski sustav	207
Slika br. 27: Proces A3: Arhivsko skladištenje mrežne građe	211
Slika br. 28: Proces A4: Upravljanje metapodacima	214
Slika br. 29: Proces A5: Korisnički pristup i Proces A6: Pakiranje i dostava	219
Slika br. 30: Proces A7: Administriranje arhivskoga sustava	220
Slika br. 31: Proces A8: Dugoročna pohrana	221
Slika br. 32: Sučelje za definiciju koda u poljima	226
Slika br. 33: Sučelje za definiciju tipova polja	227
Slika br. 34: Sučelje za definiciju strukture bibliografskih formata	228
Slika br. 35: Sučelje za oblikovanje maske za unos podataka	229
Slika br. 36: Sučelje za opis digitalnih objekata - unos trajnih identifikatora	230
Slika br. 37 : Sučelje za opis digitalnih objekata - unos bibliografskoga opisa	231
Slika br. 38: Sučelje za opis digitalnih objekata - maska za popravljavanje zapisa	232
Slika br. 39: Sučelje za grupiranje mrežne građe u sadržajne jedinice s jednim bibliografskim opisom	233

Popis dodataka

Dodatak br. 1: Osnutak strukture formata depozitarne baze mrežne građe	267
Dodatak br. 2: Prijedlog za tvorbu trajnih identifikatora URN-a u Narodnoj i univerzitetskoj knjižnici	279
Dodatak br. 3: Obrazac za slanje/prihvatanje elektroničke građe	284

UPOTRIJEBLJENE KRATICE

AACR - *Anglo-American Cataloguing Rules*
AIC - *Archival Information Collection*
AIP - *Archival Information Package*
AIU - *Archival Information Unit*
AOLAP - *Austrian Online Archive Processing (projekt)*
ARNES - *Academic Research Network of Slovenia*
ASCII - *American Standard Code for Information Interchange*
ASP - *Active Server Pages*
CCSDS - *Consultative Committee for Space Data Systems*
CEDARS - *CURL Exemplars in Digital Archives*
CENL - *Conference of Directors of European National Libraries*
CGI - *Common Gateways Interface*
COBIB - *Bibliografska baza u COBISS-u*
COBISS - *Co-operative Online Bibliographic System and Services*
COMARC - *Slovenski MARC*
CSC - *Center for Scientific Computing*
CURL - *The consortium of University Research Libraries*
DACHS - *Digital Archive for Chinese Studies*
DAMP - *Digital Archive for Web Publications*
DARSI - *Digitalni ARhiv Slovenije*
DINI - *Deutsche Initiative für Netzwerkinformation*
DIP - *Dissemination information package*
DLT - *Digital Linear Tape*
DNEP - *Depot van Nederlandse Electronische Publicaties*
DNS - *Domain name system*
DSEP - *Deposit System for Electronic Publications*
DTD - *Dana Type Definition*
ECDL - *European Conference on Research and Advanced Technology for Digital Libraries*
ECUP - *The European Copyright User Platform*
ELAG - *European Library Automation Group*
FAIFE - *IFLA Committee on Free Access to Information and Freedom of Expression*
FAQ - *Frequently Asked Questions*
FEP - *Federation of European Publishers*
FTP - *File Transfer Protocol*
FRANAR - *Functional Requirements for Authority Records*
FRBR - *Functional Requirements for Bibliographic Records*
HTML - *Hypertext Mark-up Language*
HTTP - *Hypertext Transfer Protocol*
HTTPS - *HyperText Transfer Protocol Secure*
IFLA - *International Federation of Library Associations and Institutions*
IIPC - *International Internet Preservation Consortium*
IJS - *Institut "Jožef Stefan"*
IP - *Internet Protocol*
ISBD(CR) - *International Standard Bibliographic Description for serials and other Continuing Resources*

ISBD(ER) - *International Standard Bibliographic Description for Electronic Resources*
ISBD(S) - *International Standard Bibliographic Description for Serials*
ISO - *International Standards Organization*
IWAW - *International Web Archiving Workshop*
JHOVE - *(projekt)*
JISC - *Joint Information Systems Committee, United Kingdom*
KB - *Koninklijke Bibliotheek* = nizozemska Kraljevska knjižnica
KOPAL - *Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen* = *Co-operative Development of a Long-Term Digital Information Archive* (projekt)
LMER - *Long-term preservation metadata for electronic resources*
MADS - *Metadata Authority Description Schema*
MARC - *MAchine-Readable Cataloging*
METS - *Metadata Encoding & Transmission Standard*
MINERVA - *Mapping the Internet: the Electronic Resources Virtual Archive*
MOA - *Making of America* (projekt)
MODS - *Metadata Object Description Schema*
NBN - *National Bibliography Number*
NDL - *National Diet Library, Japan*
NEDLIB - *Networked European Deposit Library*
NESTOR - *(projekt)*
NNTP - *Network News Transfer Protocol*
NORDINFO - *Nordic Council for Scientific Information*
NSK - *Narodna i sveučilišna knjižnica, Hrvatska*
NUK - *Narodna in univerzitetna knjižnica* = *Narodna i sveučilišna knjižnica, Slovenija*
NWA - *Nordic Web Archive*
OAI-PMH - *Open Archives Initiative Protocol for Metadata Harvesting*
OAIS - *Open Archival Information System*
OCLC - *Online Computer Library Center*
PANDAS - *PANDORA Digital Archive*
PANDORA - *Preserving and Accessing Networked Documentary Resources of Australia*
PDF - *Portable Document Format*
PDI - *Preservation Description Information*
PREMIS - *Preservation Metadata: Implementation Strategies*(projekt)
RFC - *Request for Comments*
RI - *Representation Information*
RIS - *Raba Interneta v Sloveniji* (projekt)
RLG - *Research Libraries Group*
SGML - *Standard Generalized Mark-up Language*
SIP - *Submission Information Package*
SNMP - *Simple Network Management Protocol*
SRCE - *Sveučilišni računski centar Sveučilišta u Zagrebu*
SSL - *Secure Socket Layer*
TAR - *Tape Archive Format*
TCP - *Transmission Control Protocol*
TOM - *Typed Object Model*
UKWAC - *UK Web Archiving Consortium*
UNIMARC - *Univerzalni MARC*

URI - *Universal Resource Identification*

URL - *Unifrom Resource Locator* = jedinstvena adresa građe

URN - *Uniform Resource Name*

UVC - *Universal Virtual Computer*

WARC - *Web ARCive file format*

WARP - *Web Archiving Project* (projekt, Japan)

WCP - *Web Characterization Project*

WEA - *Web aRchive Access*

WOM - *World of Mouth*

XML - *eXtensible Mark-up Language*

DODATAK BR. 1

OSNUTAK STRUKTURE FORMATA DEPOZITARNE BAZE MREŽNE GRADE

Pripremila: Alenka Kavčič-Čolić

Pregledale: Irena Kavčič i Špela Zupanc

Pored opisane strukture bi se morali automatsko upisivati slijedeći podatci:

- ime kreatorja
- datum kreiranja
- ime redaktora
- datum redakcije

OZNAKA POLJA I POTPOLJA	IME POLJA I POTPOLJA ³⁷³ (Kratice "m" označava koliko mjesta može zauzeti podatak u podpolju) (numerička polja su označena sa N, ostala su alfanumerična)	PONOVLJIVOST POLJA I POTPOLJA P=ponovljivo N=neponovljivo
0XX - BLOK ZA IDENTIFIKACIJU		
001	Identifikator zapisa	N
\$a	Stanje zapisa (1m)	N
\$b	Vrsta zapisa (1m)	N
\$c	Bibliografska razina (1m)	N
\$d	Hierarhijska razina (1m)	N
\$t	Tipologija dokumenta/djela (unos koda tipa br. NN.NN)	N
010	ISBN	P
\$a	Broj ISBN (13m)	N
\$b	Objašnjenje	N
\$z	Pogrešan broj	P
011	ISSN (9m)	P
\$e	Ispravan broj ISSN	N
\$c	Lokalna ser. publikacija	N
\$f	Neverificirani ISSN	N
\$a	ISSN kod članka	N
\$y	Poništen broj ISSN od strane ISDS središta	P
\$z	Pogrešan broj ISSN	P
\$s	ISSN serije, podserije, priloge za pov. podzb. s člankom	N
013	ISMN	P
\$a	Broj ISMN (13m)	N
\$b	Objašnjenje ako zapis sadrži više od jednog ISSN-a	N

³⁷³ Za prijevod u hrvatski jezik je korišten "Priručnik za UNIMARC : bibliografski format / prevela i priredila Mirna Willer. - 2. hrv. Izd. - Zagreb: Nacionalna i sveučilišna knjižnica : Hrvatsko knjižničarsko društvo, 1999. - (Izdanja Nacionalne i sveučilišne knjižnice. Priručnici ; knj. 1)

	\$Z	Pogrešan broj ISMN	P
020		Broj nacionalne bibliografije	P
	\$a	Kod zemlje (3m)	N
	\$b	Broj za ustanovu koja izdaje bibliografiju	N
	\$Z	Pogrešan broj pripisan zapisu	P
021		Broj obveznog primjerka	P
	\$a	Kod zemlje (3m)	N
	\$b	Broj	N
	\$Z	Pogrešan broj	P
1XX – BLOK KODIRANIH INFORMACIJA			
100		Opći podatci za obradu	N
	\$b	Vrsta godine izdavanja (1m)	N
	\$c	Godina izdavanja 1 (LLLL)	N
	\$d	Godina izdavanja 2 (LLLL)	N
	\$e	Kod za namjenu (1m)	N
	\$f	Kod službene publikacije (1m)	N
	\$g	Kod modificiranog zapisa (1m)	N
	\$h	Kod jezika katalogiziranja (3m)	N
	\$i	Kod za transliteraciju (1m)	N
	\$l	Pismo stvarnog naslova (2m)	N
101		Jezik jedinice	N
1.ind – prijevod 0=izvirni jez. 1=prijevod 2=sadrži prijevode 2. ind. = #			
	\$a	Jezik teksta, zvučne snimke itd. (3m)	P
	\$b	Jezik posrednog teksta kada djelo nije prevedeno s izvornika (3m)	P
	\$c	Jezik izvornog djela kada se radi o prijevodu (3m)	N
	\$d	Jezik sažetka (3m)	P
	\$e	Jezik stranice sadržaja (3m)	P
	\$f	Jezik naslovne stranice, ako se razlikuje od jezika teksta (3m)	P
	\$g	Jezik glavnog stvarnog naslova, ako nije prvi jezik teksta, zvučne snimke itd. (3m)	P
	\$h	Jezik libreta i sl., jezik tiskane muzikalije bilo kao popratna građa ili tiskano s jedinicom (3m)	P
	\$i	Jezik popratne građe (3m)	P
	\$j	Jezik podnaslova (3m)	P
102		Zemlja izdavanja ili proizvodnje	N
	\$a	Kod zemlje u kojoj je publikacija proizvedena (3m)	P
110		Polje kodiranih podataka: Serijske publikacije	N
	\$a	Vrsta serijske publikacije (1m)	N
	\$b	Učestalost izlaženja (1m)	N
	\$c	Redovitost (1m)	N
	\$d	Koda za vrstu građe (1m)	N
115		Polje kodiranih podataka: Vizualne projekcije, videosnimke i film	P
	\$a	Vrsta građe (1m)	N

	\$b	Dužina	N
	\$c	Oznaka boje (1m)	N
	\$d	Medij za zvuk (1m)	N
	\$e	Zvučni medij (1m)	N
	\$h	Tehnika - videosnimka, film (1m)	N
	\$k	Oblik puštanja u promet - videosnimka (1m)	N
	\$l	Format prikazivanja - videosnimka (1m)	N
116		Polje kodiranih podataka: Grafika	P
	\$a	Posebna oznaka građe (1m)	N
	\$b	Građa primarne podloge (1m)	N
	\$c	Građa sekundarne podloge (1m)	N
	\$d	Boja (1m)	N
	\$e	Tehnike (crteži, slike) (2m)	N
	\$f	Tehnike (otisci) (2m)	N
	\$g	Oznaka namjene (2m)	N
120		Polje kodiranih podataka: Kartografska građa - Opći podatci	N
	\$a	Oznaka za boju (1m)	N
	\$b	Oznaka za kazalo (1m)	N
	\$c	Oznaka za popratni tekst (1m)	N
	\$d	Kodovi za reljef (1m)	N
	\$e	Projekcija zemljopisne karte (2m)	N
	\$f	Početni meridijan (2m)	N
121		Polje kodiranih podataka: Kartografska građa - Opći podatci - Materijalna obilježja	N
	\$a	Dimenzije (1m)	N
	\$b	Primarni kartografska slika (1m)	N
	\$c	Fizički medij (2m)	N
	\$d	Tehnika izrade (1m)	N
	\$e	Oblik reprodukcije (1m)	N
	\$f	Geodetsko podešavanje (1m)	N
	\$g	Materijalni oblik publikacije (1m)	N
122		Polje kodiranih podataka: Vremensko razdoblje sadržaja jedinice	P
	\$a	Vremensko razdoblje	N
123		Polje kodiranih podataka: Kartografska građa - Mjerilo i koordinate	N
	\$a	Vrsta mjerila (1m)	P
	\$b	Stalan odnos linearnoga vodoravnog mjerila	P
	\$c	Stalan odnos linearnoga okomitog mjerila	P
124		Polje kodiranih podataka: Kartografska građa - Analiza posebne oznake građe	N
	\$b	Oblik kartografske jedinice (1m)	P
135		Polje kodiranih podataka: elektronička građa	P
	\$a	Vrsta elektroničke građe (1m)	N
	\$b	Posebna oznaka građe	N
	\$c	Boja	N
	\$d	Mjere	N
	\$e	Zvuk	N

\$f	Bitna dubina slike	N
\$g	Broj datotečnih formata	N
\$h	Pomagala za kvalitetno namještenje	N
\$i	Prethodnik/izvor	N
\$j	Razina kompresije	N
\$k	Kvaliteta preformatiranja	N
2XX – BLOK GLAVNOG OPISA		
200 1. indikator: 0 =naslov nije važan 1=naslov je važan 2. ind. = #	Stvarni naslovi i podatci o odgovornosti	N
\$a	Glavni stvarni naslov	P
\$b	Opća oznaka građe	P
\$c	Glavni stvarni naslov drugog autora	P
\$d	Usporedni stvarni naslov	P
\$e	Podnaslov	P
\$f	Prvi podatak o odgovornosti	P
\$g	Idući podatci o odgovornosti	P
\$h	Oznaka dijela	P
\$i	Stvarni naslov dijela	P
205	Izdanje	N
\$a	Podatak o izdanju	N
\$b	Dodatni podatak o izdanju	P
\$d	Usporedni podatak o izdanju	P
\$f	Podatci o odgovornosti koji se odnose na izdanje	P
\$g	Idući podatci o odgovornosti	P
206	Skupina posebne građe: Kartografska građa - Matematički podatci	P
\$a	Matematički podatci	N
207 1. ind. = # 2. indikator: 0= podatci formatirani, 1= podatci nisu formatirani	Skupina posebne građe: Serijske publikacije - Numerički podatci	N
\$a	Numerički podatci	P
208	Skupina posebne građe: Posebni podatci za tiskane muzikalije	N
\$a	Posebni podatci za tiskane muzikalije	N
210	Izdavanje, raspačavanje itd.	N
\$a	Mjesto izdavanja i/ili raspačavanja	P
\$b	Adresa nakladnika i/ili raspačavatelja	P
\$c	Nakladnik i/ili raspačavatelj	P
\$d	Godina izdavanja i/ili raspačavanja	N
215	Materijalni opis	N
\$a	Posebna oznaka građe i opseg	N
\$c	Druge materijalne pojedinosti	N
\$d	Dimenzije	N
\$e	Podatak o popratnoj građi	P

	\$g	Numerički podatci - treća razina	N
	\$i	Numerički podatci – druga razina (vol.)	N
	\$h	Numerički podatci – prva razina (broj)	N
	\$k	Kronologija (godina)	N
	\$o	Alternativna paginacija	N
	\$p	Alternativni numerički podatci – III	N
	\$q	Alternativni numerički podatci - II	N
	\$r	Alternativni numerički podatci – I	N
	\$s	Alternativna kronologija (godina)	N
225		Nakladnička cjelina	P
	\$a	Skupni stvarni naslov	N
	\$d	Usporedni skupni stvarni naslov	P
	\$e	Podnaslov nakladničke cjeline	P
	\$f	Podatci o odgovornosti koji se odnose na nakladničku cjelinu	P
	\$h	Oznaka kola (razdjela)	P
	\$i	Stvarni naslov kola (razdjela)	P
	\$v	Numeracija u nakladničkoj cjelini	P
	\$x	ISSN	P
230		Skupina posebne građe: Svojstva elektroničke građe	P
	\$a	Oznaka in opseg datoteke	N
3XX – BLOK NAPOMENA			
300		Opća napomena	P
	\$a	Tekst napomene	N
320		Napomena o bibliografiji/kazalima/sažecima što ih publikacija sadrži	P
	\$a	Tekst napomene	N
330		Kratki sadržaj ili sinopsis	P
	\$z	Jezik napomene (3m)	N
	\$a	Tekst napomene	N
	\$f	Autor sadržaja ili sinopsisa	N
4XX – BLOK ZA POVEZIVANJE KATALOŽNIH JEDINICA			
410		Nakladnička cjelina (niz)	P
	\$a	Stvarni naslov	N
	\$x	ISSN (nakladničke cijelne čije je kolo)	N
411		Kolo	P
	\$a	Stvarni naslov	N
	\$x	ISSN (kola)	N
421		Prilog	P
	\$a	Stvarni naslov	N
	\$x	ISSN (prilozi)	N
422		Matična publikacija	P
	\$a	Stvarni naslov	N
	\$x	ISSN	N
430		Nastavak publikacije	P
	\$a	Stvarni naslov	N

	\$x	ISSN	N
431		Djelomice je nastavak publikacije	P
	\$a	Stvarni naslov	N
	\$x	ISSN	N
434		Preuzima publikaciju	P
	\$a	Stvarni naslov	N
	\$x	ISSN	N
435		Djelomice preuzima publikaciju	P
	\$a	Stvarni naslov	N
	\$x	ISSN	N
436		Nastaje spajanjem publikacija	P
	\$a	Stvarni naslov	N
	\$x	ISSN	N
440		Nastavlja se kao	P
	\$a	Stvarni naslov	N
	\$x	ISSN	N
441		Djelomice se nastavlja kao	P
	\$a	Stvarni naslov	N
	\$x	ISSN	N
444		Preuzima je	P
	\$a	Stvarni naslov	N
	\$x	ISSN	N
445		Djelomice je preuzima	P
	\$a	Stvarni naslov	N
	\$x	ISSN	N
446		Dijeli se na ..., ... i na ...	P
	\$a	Stvarni naslov	N
	\$x	ISSN	N
447		Spaka se sa ... i ...te nastaje ...	P
	\$a	Stvarni naslov	N
	\$x	ISSN	N
448		Vraća se na	P
	\$a	Stvarni naslov	N
	\$x	ISSN	N
452		Druga izdanja u drugome mediju	P
	\$l	Povezivanje sa drugim izdanjima	N
	\$a		N
	\$f		N
453		Prijevod	P
	\$a	Stvarni naslov	N
	\$x	ISSN (drugog izdanja)	N
454		Izvornik	N
	\$a	Stvarni naslov	N
	\$x	ISSN (izvornog izdanja)	N
463		Dio	P

	\$l	Identifikacijski broj zapisa	N
	\$a	Stvarni naslov	N
	\$f	Numerički podatci (vol., br.)	N
464		Analitički dio	P
	\$l	Identifikacijski broj zapisa	N
	\$t		N
	\$v		N
488		Dijelo povezivano na drugi način	P
	\$a	Stvarni naslov	N
	\$x	ISSN	N
5XX – BLOK SRODNIH NASLOVA			
500		Jedinstveni stvarni naslov	P
	\$a	Jedinstveni stvarni naslov	P
	\$h	Oznaka dijela	P
	\$i	Stvarni naslov dijela	P
	\$m	Jezik (3m)	N
503		Jedinstvena uobičajena odrednica	N
	\$a	Formalna odrednica	N
	\$j	Godina	N
517		Ostali sporedni stvarni naslovi	P
	\$a	Sporedni stvarni naslov	N
530		Ključni naslov (Serijske publikacije)	N
	\$a	Ključni naslov	N
	\$b	Dopuna	N
531		Skraćeni naslov (Serijske publikacije)	N
	\$a	Skraćeni naslov	N
	\$b	Dopuna	N
532		Prošireni stvarni naslov	P
	\$a	Prošireni stvarni naslov	N
540		Dodatni naslov koji daje katalogizator	P
	\$a	Dodatni naslov	N
6XX – BLOK SADRŽAJNE ANALIZE			
600		Osobna predmetna odrednica	P
	\$a	Prvi element unosa	P
	\$b	Dio imena koji nije prvi element unosa	P
	\$c	Osobna obilježja osim datuma	P
	\$d	Rimski brojevi	P
	\$f	Datumi	P
601		Korporativna predmetna odrednica	P
	\$a	Ime korporacije (Prvi element unosa)	P
	\$b	Podjela (ili naziv tijela ako se unosi pod sjedištem)	P
	\$c	Dodatak nazivu ili dopuna	P
	\$d	Redni broj sastanka	P
	\$e	Mjesto sastanka	P
	\$f	Godina održavanja sastanka	P

	\$g	Premetnuti element	P
	\$h	Dio naziva koji nije prvi element unosa ni premetnuti element	P
602		Obiteljska predmetna odrednica	P
	\$a	Prvi element unosa	P
	\$f	Datumi	P
605		Stvarni naslov kao predmetna odrednica	P
	\$a	Prvi element unosa	P
606		Tematska predmetna odrednica	P
	\$a	Prvi element unosa	P
607		Zemlopisna predmetna odrednica	P
	\$a	Prvi element unosa	P
609		Oblik, žanr i/ili materijalna obilježja kao predmetna odrednica	P
	\$a	Prvi element unosa	P
610		Slobodno oblikovane predmetnice	P
	\$a	Prvi element unosa	P
	\$z	Jezik (3m)	N
675		UDK	P
	\$a	Broj	N
	\$c	Kod	N
	\$s	Statistika	N
	\$v	Izdanje	N
	\$z	Jezik izdanja	N
7XX - BLOK PODATAKA O ODGOVORNOSTI			
700 2. ind. =#		Osobno ime – primarna odgovornost	N
	\$a	Prvi element unosa	N
	\$b	Dio imena koji nije prvi element unosa	N
	\$c	Osobna obilježja osim datuma	P
	\$d	Rimski brojevi	N
	\$f	Datumi	N
	\$8	Afilijacija/adresa	P
	\$3	Broj normativnog zapisa	N
	\$4	Kod koji se odnosi na vrstu autorstva	P
701 1. ind. =# 2. ind. =#		Osobno ime – alternativna odgovornost	P
	\$a	Prvi element unosa	N
	\$b	Dio imena koji nije prvi element unosa	N
	\$c	Osobna obilježja osim datuma	P
	\$d	Rimski brojevi	N
	\$f	Datumi	N
	\$8	Afilijacija/adresa	P
	\$3	Broj normativnog zapisa	N
	\$4	Kod koji se odnosi na vrstu autorstva	P
702 1. ind. =# 2. ind. =#		Osobno ime – sekundarna odgovornost	P

	\$a	Prvi element unosa	N
	\$b	Dio imena koji nije prvi element unosa	N
	\$c	Osobna obilježja osim datuma	P
	\$d	Rimski brojevi	N
	\$f	Datumi	N
	\$8	Afilijacija/adresa	P
	\$3	Broj normativnog zapisa	N
	\$4	Kod koji se odnosi na vrstu autorstva	P
710 1. ind. =# 2. ind. =#		Naziv korporativnog tijela – primarna odgovornost	N
	\$a	Prvi element unosa	N
	\$b	Podjela (ili naziv tijela ako se unosi pod sjedištem)	P
	\$c	Dodatak nazivu ili dopuna	P
	\$d	Redni broj sastanka	N
	\$e	Mjesto sastanka	P
	\$f	Godina održavanja sastanka	N
	\$g	Premetnuti element	N
	\$h	Dio naziva koji nije prvi element unosa, ni premetnuti element	P
	\$3	Broj normativnog zapisa	N
	\$4	Kod koji se odnosi na vrstu autorstva	P
711 1. ind. =# 2. ind. =#		Naziv korporativnog tijela – alternativna odgovornost	P
	\$a	Prvi element unosa	N
	\$b	Podjela (ili naziv tijela ako se unosi pod sjedištem)	P
	\$c	Dodatak nazivu ili dopuna	P
	\$d	Redni broj sastanka	N
	\$e	Mjesto sastanka	P
	\$f	Godina održavanja sastanka	N
	\$g	Premetnuti element	N
	\$h	Dio naziva koji nije prvi element unosa, ni premetnuti element	P
	\$3	Broj normativnog zapisa	N
	\$4	Kod koji se odnosi na vrstu autorstva	P
712 1. ind. =# 2. ind. =#		Naziv korporativnog tijela – sekundarna odgovornost	P
	\$a	Prvi element unosa	N
	\$b	Podjela (ili naziv tijela ako se unosi pod sjedištem)	P
	\$c	Dodatak nazivu ili dopuna	P
	\$d	Redni broj sastanka	N
	\$e	Mjesto sastanka	P
	\$f	Godina održavanja sastanka	N
	\$g	Premetnuti element	N
	\$h	Dio naziva koji nije prvi element unosa, ni premetnuti element	P
	\$3	Broj normativnog zapisa	N
	\$4	Kod koji se odnosi na vrstu autorstva	P

8XX – BLOK MEĐUNARODNE UPOTREBE		
850	Ustanova u kojoj se nalazi građa	P
\$a	Kod ustanove (p)	N
856 Indikator 1. način pristupa 2. odnos	Elektronička lokacija i pristup	P
\$a	Naziv domaćina	P
\$b	Broj pristupa	P
\$c	Informacija o kompresiji	P
\$d	Put	P
\$e	Datum i sat pristupa	P
\$f	Elektronički naziv	P
\$g	Univerzalni naziv građe (URN)	P
\$h	Procesor prema kojem se upućuje zahtjev	N
\$i	Uputa	P
\$j	Bitovi u sekundi	N
\$k	Lozinka	N
\$l	Prijava s udaljene lokacije	N
\$m	Kontakt za pomoć pri pristupu	P
\$n	Naziv lokacije domaćina u potpolju \$a	N
\$o	Operacijski sustav	N
\$p	Port	N
\$q	Vrsta elektroničkog formata	N
\$r	Parametri komunikacije	N
\$s	Veličina datoteke	P
\$t	Emulacija terminala	P
\$u	Jedinstveni lokator građe (URL)	N
\$v	Vrijeme u koje je način pristupa dostupan	P
\$w	Kontrolni broj zapisa	P
\$x	Napomena koja nije namijenjena korisnicima	P
\$y	Način pristupa	N
\$z	Napomena za korisnike	P
886	Podatci koji nisu bili preneseni iz izvornog formata	P
\$a	Oznaka u izvornom formatu	N
\$b	Indikatori i potpolja u izvornom formatu	N
\$2	Kod sustava	N
9XX – BLOK ZA NACIONALNU UPOTREBU		
900 2. ind. =#	Osobno ime – primarna odgovornost	N
\$a	Prvi element unosa	N
\$b	Dio imena koji nije prvi element unosa	N
\$c	Osobna obilježja osim datuma	P
\$d	Rimski brojevi	N
\$f	Datumi	N
\$3	Broj normativnog zapisa	N

901 1. ind. =# 2. ind. =#	Osobno ime – alternativna odgovornost	P
\$a	Prvi element unosa	N
\$b	Dio imena koji nije prvi element unosa	N
\$c	Osobna obilježja osim datuma	P
\$d	Rimski brojevi	N
\$f	Datumi	N
\$3	Broj normativnog zapisa	N
902 1. ind. =# 2. ind. =#	Osobno ime – sekundarna odgovornost	P
\$a	Prvi element unosa	N
\$b	Dio imena koji nije prvi element unosa	N
\$c	Osobna obilježja osim datuma	P
\$d	Rimski brojevi	N
\$f	Datumi	N
\$3	Broj normativnog zapisa	N
910 1. ind. =# 2. ind. =#	Ime korporacije – primarna intelektualna odgovornost	N
\$a	Prvi element unosa	N
\$b	Podjela (ili naziv tijela ako se unosi pod sjedištem)	P
\$c	Dodatak nazivu ili dopuna	P
\$d	Redni broj sastanka	N
\$e	Mjesto sastanka	P
\$f	Godina održavanja sastanka	N
\$g	Premetnuti element	N
\$h	Dio naziva koji nije prvi element unosa, ni premetnuti element	P
\$3	Broj normativnog zapisa	N
911 1. ind. =# 2. ind. =#	Ime korporacije – alternativna intelektualna odgovornost	P
\$a	Prvi element vnosa	N
\$b	Podjela (ili naziv tijela ako se unosi pod sjedištem)	P
\$c	Dodatak nazivu ili dopuna	P
\$d	Redni broj sastanka	N
\$e	Mjesto sastanka	P
\$f	Godina održavanja sastanka	N
\$g	Premetnuti element	N
\$h	Dio naziva koji nije prvi element unosa, ni premetnuti element	P
\$3	Broj normativnog zapisa	N
912 1. ind. =# 2. ind. =#	Ime korporacije – sekundarna intelektualna odgovornost	P
\$a	Prvi element vnosa	N
\$b	Podjela (ili naziv tijela ako se unosi pod sjedištem)	P
\$c	Dodatak nazivu ili dopuna	P
\$d	Redni broj sastanka	N

	\$e	Mjesto sastanka	P
	\$f	Godina održavanja sastanka	N
	\$g	Premetnuti element	N
	\$h	Dio naziva koji nije prvi element unosa, ni premetnuti element	P
	\$3	Broj normativnog zapisa	N
950		Podatci o promjenama elektroničke građe	P
	\$a	Datum promjena (LLLLMMDD)	N
	\$b	Opis promjene	N
	\$c	Autor promjene	N
951		Podatci o autorskim pravima	P
	\$a	Vlasnik autorskih prava	N
	\$b	Broj ugovora	N
	\$c	Vrijeme zaštite	N
	\$d	Uvjeti pristupa	N
	\$e	Dozvoljene intervencije u elektroničkoj građi	N
	\$f	Osobe kojima je pristup dozvoljen	N
	\$g	Drugi važni poaci koji nisu obuhvaćeni od \$a-\$g	
952		Informacije o vjerodostojnosti građe (Fixity)	P
	\$a	<i>Checksum</i>	N
	\$d	Drugo	P
953		Upute za upotrebu	P
	\$a	Opis	N
954		Kompajleri i interpreteri za različite progr. jezike	P
	\$a	Ime	
	\$b	Verzija	N
	\$c	Upute	
955		Drugi podatci o skladištenju	P
	\$a	Opis	N
992		Polje za lokalne potrebe	N
	\$b	Oznaka zapisa	N

DODATAK BR. 2

PRIJEDLOG ZA TVORBU TRAJNIH IDENTIFIKATORA URN-a
U NARODNOJ I UNIVERZITETSKOJ KNJIŽNICI

Trajni identifikatori potrebni su nam da bismo jednoznačno identificirali elektroničku građu te je pronašli u svom digitalnom arhivu bez obzira na poslužitelja na kome je pohranjena ta građa. Postoji više vrsta identifikatora, naprimjer PURL, DOI, ARK, URN itd. Iskustva nekih evropskih nacionalnih knjižnica pokazuju da je optimalna upotreba URN-a jer su besplatni, jednostavno ih je kreirati, a njihovu strukturu možemo sami odrediti. Za razliku od ostalih identifikatora, URN (*Uniform Resource Name* = jedinstveni naziv građe) može sadržati i informaciju o građi, kao što su formalni bibliografski identifikatori ili signature.

Postoje dva načina tvorbe URN-a:

- umjetno na osnovi nekoga tekućega broja i šifre jedinice građe,
- na osnovi postojećih podataka (formalno dodijeljenih bibliografskih identifikatora ili koda automatskih generiranih brojeva, kao što su kontrolni zbroj (*checksum* ili MD5) ili signature).

Prednost umjetno stvorenih URN-a je ta da su međusobno slični i kraći te je njima lakše upravljati. Slijede strukturu:

URN:NBN:SI:NUK:šifrant_tipa_građe/umjetno_stvoren_broj

Primjer:: URN:NBN:SI:NUK:a19844443

Nedostatak ovog tipa URN-a je taj da se moraju stvarati centralizirano pomoću računalne programske aplikacije, a za njihovo je vođenje potreban registar. Osim toga, iz URN-a nije moguće pročitati o kakvoj građi se radi, što u velikoj mjeri otežava kontrolu duplikata. Zato je druga metoda za stvaranje URN-a mnogo prikladnija i u nastavku je dan prijedlog tvorbe mogućih struktura.

1. Tvorba URN-a na osnovi postojećih formalno dodijeljenih bibliografskih identifikatora (ISBN, ISSN, ISIC ...)

MONOGRAFSKE ELEKTRONIČKE PUBLIKACIJE

URN:	NBN:SI:	NUK:	ISBN:	961-6162-83-7	<pdf>	A	0.1
Tip identifikatora	Uključivanje u slovensku nacionalnu bibliografiju	Institucija	Tip postojećeg bibliografskog identifikatora	Broj identifikatora	Format datoteke	Verzija ili podatci o kopiji	Izdanje ili verzija publikacije
		CTK UKM EFLJU	ISSN	0015-6914	<doc> <tiff>	A = originalna datoteka B = rezultat konverzije originalne datoteke D = digitalna kopija originala I = prikupljeno na internetu	0.1 0.2 ...

Primjer:

URN:NBN:SI:NUK:ISBN:961-6162-83-7<PDF>A0.1

(URN publikacije: ISBN : Priročnik za uporabnike. Prevedla in priredila Alenka Kanič; Elektronsko verzijo pripravil Ivan Kanič. - Ljubljana : NUK, 2002. Prevod dela: ISBN users' manual. - Fourth, revised, international ed.. - Berlin : International ISBN Agency. - ISBN 3-88053-075-0.)

Trebalo bi definirati protokol za rješavanje sljedećih problema:

- Manje promjene u tekstovima nemaju uvijek utjecaj na promjenu ISBN-a u elektroničkom okruženju. Kako ćemo onda označiti "različito izdanje"? Postoje tri mogućnosti:
 - o da se iza podatka o izdanju datoteke doda datum preuzimanja,
 - o da se doda još jedan šifrant koji označava izdanje u arhivu,
 - o da se doda *checksum* ili jedinstveni identifikacijski broj MD5.
 Vjerojatno je drugo rješenje najbolje jer se kod prvog može dogoditi da se datum preuzimanja zamijeni s datumom objave, a treće je rješenje predugo.
- Postoje zbornici koji se izdaju jednom godišnje, a istovremeno se obrađuju i kao knjige i kao serijske publikacije, što bi moglo uvesti zbrku kod tvorbe URN-a. U tim je slučajevima bolje URN-e tvoriti na osnovi ISBN-a, a ne na osnovi ISSN-a.

SERIJSKE PUBLIKACIJE

URN:	NBN:SI:	NUK:	ISSN:	1580-5352	<20060214>	<html>	I
Tip identifikatora	Uključivanje u slovensku nacionalnu bibliografiju	Institucija	Tip postojećeg bibliografskog identifikatora	Broj identifikatora	Datum objave u formatu ISO ili vol; br.	Format datoteke	Verzija ili podatci o kopiji
		CTK UKM EFLJU	ISBN	0015-6914	20051212 <34;4>	<doc> <tiff>	A = originalna datoteka B = rezultat konverzije originalne datoteke D = digitalna kopija originala I = prikupljeno na internetu

Primjer:

URN:NBN:SI:NUK:ISSN:1580-5352<20060214><html>I

(URN revije: Mladina na mreži, objavljena 14.2.2006.).

Kod serijskih publikacija treba računati na sljedeće scenarije:

- za elektroničke serijske publikacije koje su objavljene samo na mreži i mijenjaju se svakodnevno ili svakog tjedna, datum objave bit će jednak datumu prikupljanja u formatu ISO; ako izlaze jednom mjesečno, onda ćemo zadnji broj u datumu zamijeniti nulama (20010200);

- ako je elektronička revija koju slovenska organizacija objavljuje više puta godišnje dostupna na mreži, onda možemo koristiti dvije varijante: upišemo datum objave u formatu ISO ili upišemo godinu i broj jedinice u uglatim zagrada ("<>") - tada oba podatka odvajamo točkom sa zarezom (","); takvu oznaku možemo koristiti kod digitaliziranih serijskih publikacija ili kod elektroničkih izvoda tiskanih serijskih publikacija na mreži.

ČLANCI I REFERATI SA STRUČNIH SKUPOVA

Postupak pri tvorbi URN-a za članke sličan je postupku pri tvorbi URN-a serijskih elektroničkih publikacija, s jedinom razlikom što su članci dio serijske publikacije i zato je dodatno brojem stranica označen njihov opseg.

URN:	NBN:SI:	NUK:	ISSN:	1580-4070	<14;1>	Str1-2	<html>	I
Tip identifikatora	Uključivanje u slovensku nacionalnu bibliografiju	Institucija	Tip postojećeg bibliografskog identifikatora	Broj identifikatora	Datum objave u formatu ISO ili vol; br.	Opseg članka	Format datoteke	Verzija ili podatci o kopiji
			ISSN, ISBD, CODEN_...	0015-6914	<20051212> <34;4> <2005;12>	Str563-617	<doc> <tiff>	A = originalna datoteka B = rezultat konverzije originalne datoteke D = digitalna kopija originala I = prikupljeno na internetu

Primjer:

URN:NBN:SI:NUK:ISSN:1580-4070<14;1>Str1-2<html>I

Većina elektroničkih serijskih publikacija nema paginacije, u tom slučaju dodajemo URL iz kojeg je članak bio prikupljen.

Primjer: Sever, Jani: Izboljševalec sveta. U: Mladina, 14.2.2006, URL: http://www.mladina.si/tehdnik/200606/clanek/uvo-uvodnik--jani_sever/.

URN:	NBN:SI:	NUK:	ISSN:	1580-4070	<20060214>	<html>	I:	URL/ MD5
Tip identifikatora	Uključivanje u slovensku nacionalnu bibliografiju	Institucija	Tip postojećeg bibliografskog identifikatora	Broj identifikatora	Datum objave u formatu ISO ili vol; br.	Format datoteke	Verzija ili podatci o kopiji	
			ISSN CODEN_...	0015-6914	<20051212> <34;4>	<doc> <tiff>	A = originalna datoteka B = rezultat konverzije originalne datoteke D = digitalna kopija originala I = prikupljeno na internetu	

Varijanta 1: dodajemo naslov URL-a:

URN:NBN:NUK:ISSN:1580-5352<20060214><html>I:http://www.mladina.si/tednik/200606/clanek/uvo-uvodnik--jani_sever/

Varijanta 2: dodajemo MD5, koji može biti kraći od naslova URL-a:

URN:NBN:NUK:ISSN:1580-5352<20060214><html>I:65EE7FC2436819EB781A841F387AF8DF

2. Tvorba URN-a za elektroničku građu koja nema bibliografskih identifikatora

Bibliografski identifikatori identificiraju tiskanu građu, a vrlo je malo elektroničke građe koja ih sadrži. Juha Hakala³⁷⁴ predlaže da se za elektroničku građu koja nema formalno dodijeljenih bibliografskih identifikatora upotrijebe brojevi nacionalne bibliografije. Za građu koja je dobivena s interneta predlaže da se upotrijebi broj iz kontrolnoga zbroja (*checksum*), koji se automatski generira. Kontrolni zbroj je zbroj bita i drugih parametra u datoteci i vrlo je mala vjerojatnost (po Hakali je ta mogućnost 2^{64} slučaj) da će se dvaput pojaviti isti broj kod dviju datoteka.

Aplikacija koju je razvio Institut "Jožef Stefan" za svaki dokument na mreži automatsko generira MD5 kod, koji je vrlo sličan kontrolnomu zbroju. Kod možemo upotrijebiti za sve elektroničke publikacije koje nemaju nikakvih identifikacijskih brojeva, i to pomoću sljedeće sintakse:

URN:NBN:SI:NUK:vg:MD5<datum>

URN:	NBN:SI:	NUK:	vg:	MD5	<html>	<datum>
Tip identifikatora	Uključivanje u slovensku nacionalnu bibliografiju	Institucija	Vrsta građe	Broj koji se automatski generira u trenutku prikupljanja pomoću robota	Tip datoteke	Datum preuzimanja u formatu ISO - generira se automatski
		CTK EFLJU	a = članak/ referat m = knjiga s = serijska publikacija d = naslovnica p = mrežno mjesto	45AE6FC1436919EB881D841E387DF8DF	Koristi se samo za pojedinačne mrežne stranice	20051212

Primjer naslovnice portala najdi.si: (URL: <http://www.najdi.si>):

URN:NBN:SI:NUK:p:00AB318AA6EC79476F59635102798845<20060214>

Primjer mrežne stranice koja je sastavni dio naslovnice na mreži:

URN:NBN:SI:NUK:p:8F5F7C1108200A26F1E14F02FE7208C8<PDF><20060214>

³⁷⁴ RFC 3188.

Reference:

RFC 2141: URN Syntax: <http://www.faqs.org/rfcs/rfc2141.html>

[<http://www.ietf.org/rfc/rfc2141.txt>](http://www.ietf.org/rfc/rfc2141.txt)

RFC 3044 - Using The ISSN (International Serial Standard Number) as URN (Uniform Resource Names) within an ISSN-URN Namespace: [<http://www.faqs.org/rfcs/rfc3044.html>](http://www.faqs.org/rfcs/rfc3044.html)

RFC 3061 - A URN Namespace of Object Identifiers: [<http://www.faqs.org/rfcs/rfc3061.html>](http://www.faqs.org/rfcs/rfc3061.html)

RFC 3187 - Using International Standard Book Numbers as Uniform Resource Names: [<http://www.faqs.org/rfcs/rfc3187.html>](http://www.faqs.org/rfcs/rfc3187.html)

RFC 3188 - Using National Bibliography Numbers as Uniform Resource Names: [<http://www.faqs.org/rfcs/rfc3188.html>](http://www.faqs.org/rfcs/rfc3188.html)

RFC 2288 - Using Existing Bibliographic Identifiers as Uniform Resource Names: [<http://www.faqs.org/rfcs/rfc2288.html>](http://www.faqs.org/rfcs/rfc2288.html)

DODATAK BR. 3

OBRAZAC ZA SLANJE/PRIHVAĆANJE ELEKTRONIČKE GRAĐE

I. Podatci o svim autorima elektroničke građe: (autori sadržaja, oblikovanja, aplikacije i slično)

1. Ime i prezime autora; ako je autor organizacija, upišite ime organizacije	2. Adresa (ulica i kućni broj, broj pošte i mjesto)	3. Kontaktni brojevi (e-pošta, fax, tel.); ako je autor organizacija, upišite ime osoba za kontakt
1.		
2.		
3.		
4.		
5.		
6.		
7.		

II. Podatci o nakladniku:

4. Ime nakladnika ili organizacije; ako je autor nakladnik, upišite samo njegovo ime	5. Adresa	6. Kontaktni brojevi (osoba, e-pošta, fax, tel.).
1.		
2.		
3.		
4.		
5.		
6.		
7.		

III. Podatci o elektroničkoj građi:

7. Naslov publikacije:

8. Ključne riječi:

9. Tip publikacije: (slovom *X* označite odgovarajući tip publikacije i dopunite podatke o serijskoj publikaciji ili članku, odnosno o sastavnom dijelu knjige)

- 9.1. monografija ____
- 9.2. serijska publikacija ____
- prva godina izlaženja: _____
 - zadnja godina izlaženja: _____
 - učestalost izlaženja: _____
 - prethodni naslov publikacije: _____
 - dodatci publikacije: _____
- 9.3. baza podataka ____
- 9.4. naslovnica ____
- 9.5. godišnji izvještaj ____
- 9.6. statistika ____
- 9.7. udžbenik ____
- 9.8. priručnik ____
- 9.9. enciklopedija ____
- 9.10. rječnik ____
- 9.11. zbornik ____
- 9.12. članak ____
- navedite podatke o serijskoj publikaciji čiji je članak sastavni dio: _____
- 9.13 referat na stručnim skupovima ili konferencijama, seminarima ...
- navedite podatke o zborniku čiji je referat sastavni dio: _____
- 9.14 sastavni dio knjige (samostalno poglavlje)
- navedite podatke o knjizi, čije je poglavlje sastavni dio: _____
- 9.15 ostalo (navedite što): _____

10 Jezik grade

- 10.1 Jezik elektroničke publikacije: _____
- 10.2 Jezik originala iz kojeg je nastala publikacija: _____

11 Karakteristike elektroničke publikacije:

- 11.1 publikacija je original ____
- 11.2 publikacija je izašla samo u elektroničkom obliku ____
- 11.3 publikacija je samo reprodukcija originala u tiskanom obliku ____
- navedite podatke o originalu: _____
- 11.4 publikacija je reprodukcija originala u tiskanom obliku, iako ima dodatnu vrijednost ____
- navedite podatke o originalu: _____
- 11.5 publikacija je novo izdanje tiskane publikacije ____
- navedite podatke o originalu: _____

11.6 publikacija je novo izdanje elektroničke publikacije ____ ○ navedite podatke o prvom izdanju: _____
11.7 publikacija je prijevod druge publikacije ____ navedite podatke o publikaciji koja je prevedena: _____
12. Datum objave: _____ (LLLLMMDD)
13. Format: 13.1 format datoteke/a (po mogućnosti šaljite datoteke u sljedećim formatima: XML/SGML, HTML, PDF, PS, RTF, TIFF, FOC, DVI, TXT) 13.2 format bibliografskoga zapisa: ○ ONIX ____ ○ DUBLIN CORE ____ ○ UNIMARC ____ ○ MARC ____ ○ DRUGO ____ 13.3 tip kompresije ako je bila uporabljena
14. Posebni tehnički zahtjevi elektroničke publikacije (u slučaju da je publikacija u nekom od formata koji nisu navedeni u točki 13.1) - Okruženje djelovanja (HW) - posebne grafičke, zvučne ili druge kartice, dodatna oprema _____ - Operacijski sustav: _____ - Veličina memorije: _____ - Brzina procesora: _____ - Kompajler: _____ - Ostali zahtjevi: _____
15. Zaštita građe (zbog dugoročne pohrane, elektronička građa mora biti bez tehničke zaštite pristupa). Nositelj autorskih prava može odrediti način pristupa građi na osnovi ugovora u prilogu. ○ Građa je zaštićena lozinkom ____ ○ Pristup je ograničen brojem IP ____ ○ Dokument je kodiran ____ ○ Druga zaštita ____
16. Identifikator: ○ URN ____ ○ PURL ____ ○ DOI ____ ○ Ostalo ____
IV. Način slanja publikacije: ○ putem e-pošte ____ ○ publikacija je dostupna na FTP-u: ____ (lozinka za pristup?) ____ ○ publikacija je dostupna na URL-u: ____ (lozinka za pristup?) ____ ○ publikaciju ćemo predati na fizičkom nosaču (disketi, CD-ROM-u, magnetnoj traci) ____ ○ ostalo ____
V. Podatci o autorskim pravima: Dvije varijante: A) napisati cijeli tekst i ponuditi autoru različite mogućnosti odabira B) napisati cijeli tekst i ostaviti prazan prostor u koji autor sam upisuje dopuštenja u vezi s autorskim pravima

Autor: Alenka Kavčič-Čolić

Naslov: Arhiviranje slovenike na internetu : metodološki pristup

Mjesto: Zagreb

Ljeto: 2007

Broj stranica: 296

Broj bibliografskih referenci: 244

Broj tablica: 8

Broj slika: 39

Mentor: prof. dr. Aleksandra Horvat

UDK: 004.738.5:005.921.1"746"(497.4)

ARHIVIRANJE SLOVENIKE NA INTERNETU : METODOLOŠKI PRISTUP

SAŽETAK

Godine 2006. u Sloveniji je prihvaćen novi Zakon o obveznom primjerku, koji obuhvaća i mrežnu građu. Narodna i sveučilišna knjižnica (NUK) po svojoj je funkciji prema tom zakonu preuzela najvišu odgovornost za prikupljanje i pohranu elektroničkih publikacija koje su na mreži i dio su slovenskoga kulturnoga naslijeđa. Zato je jednim od njezinih primarnih zadataka postala izgradnja web arhiva.

Stoga su ciljevi disertacije: naći optimalan način za prikupljanje i pohranu elektroničkih publikacija koje su objavljene na internetu, prilagoditi različite međunarodne metodologije i izraditi metodologiju za arhiviranje mreže koja bi bila primjerena za NUK te prikazati bitne aspekte arhiviranja slovenskih publikacija na mreži, koji mogu pridonijeti odlučivanju kod odabira optimalnih rješenja.

Polazeći od različitih međunarodnih iskustava i prakse, u ovom radu posebnu pozornost autorica daje izgradnji web arhiva. Pritom je analizirana mrežna građa, način na koji njezine karakteristike utječu na odabir metode i strategija njihova prikupljanja. Kvaliteta prikupljene građe ovisi o postojećim alatima za prikupljanje. Predstavljene su različite tipologije interneta, koje se temelje prije svega na sadržaju, formi ili funkcionalnosti jedinica građe, kao alternativne podjele koje donosi genologija interneta. Sagledani su i problemi bibliografske obradbe i autorskoga prava. Veliku pozornost autorica daje kriterijima odabira, koji mogu biti sadržajni, formalni ili tehnički. Vrlo značajan aspekt čine postojeće strategije za dugoročnu pohranu mrežne građe, uključujući i

metapodatke kao nosioce dragocjenih informacija za osiguravanje pristupa do mrežne građe u budućnosti.

Rezultat je istraživanja sveobuhvatni metodološki model, koji se temelji na referentnom modelu Otvorenoga arhivskoga informacijskoga sustava (OAIS - *Open Archival Information System*) i koji će biti osnova za uspostavljanje sustava za arhiviranje mrežne građe u NUK-u, a koji je moguće primijeniti i u drugim okruženjima.

Ključne riječi: web arhivi, izgradnja web arhiva, mrežna građa, prikupljanje mrežne građe, dugoročna pohrana mrežnih stranica, metode, metodologija, modeli, OAIS.

ARCHIVING SLOVENICA ON THE INTERNET: A METHODOLOGICAL APPROACH

ABSTRACT

In 2006 a new law on legal deposit was passed by the Slovenian Parliament which included web publications too. According to this law, the National and University Library (NUK) was designated as the institution responsible for the collection and long-term preservation of web publications, which are part of Slovenian cultural heritage. As consequence, the development of the web archive became one of its primary tasks.

In this regard, the aim of the thesis are the following: to find optimal methods for collecting and preserving web publications; to adjust different international methodologies and create a suitable one for NUK; and last but not least, to present important related aspects of web archiving which could contribute in the decision process concerning optimal solution.

Considering various international experiences and practices, special attention to web archive development is given. Specific characteristics of web publications and their influence on the selection of different collection methods and strategies are discussed as well as the possibilities of

the existing harvesting tools that could influence the quality of archived materials. Different internet typologies based on the content, form and functionality of web publications are presented. Interesting solutions can be found in the taxonomy, introduced by Internet genre theory. The problems of bibliographic description of web publications and copyright are described. Special emphasis is given to web selection criteria, with reference to the content, form or technical level of web publications. Existing strategies for long-term preservation, including metadata as holders of valuable information for securing access to web contents in the future, are very important.

The result of this research is a methodological model based on OAIS (Open Archival Information System) Reference model which will be used for the implementation of NUK's web archiving system. This model will be possible to implement in other environments too.

Kew Words: web archives, web archives development, web publications, web publications collection, long-term preservation, methods, methodology, models, OAIS.

ARHIVIRANJE SLOVENIKE NA INTERNETU: METODOLOŠKI PRISTOP

POVZETEK

Leta 2006 je bil v Sloveniji sprejet novi Zakon o obveznem izvodu, ki zajema tudi gradivo na medmrežju. Narodna in univerzitetna knjižnica (NUK) je zaradi svoje funkcije po tem zakonu prevzela največjo odgovornost za zbiranje in ohranjanje elektronskih publikacij, ki so objavljene na medmrežju in so del slovenske kulturne dediščine, zato je ena izmed njenih primarnih nalog postala izgradnja spletnega arhiva.

Cilji disertacije so: najti optimalni način za zbiranje in ohranjanje elektronskih publikacij, ki so bile objavljene na medmrežju; prilagoditi različne mednarodne metodologije in izdelati metodologijo za arhiviranje publikacij na medmrežju, ki bi bila primerna za NUK; ter predstaviti pomembne aspekte arhiviranja slovenskih publikacij na medmrežju, ki bi lahko prispevali pri izbiri najboljših rešitev.

Na podlagi različnih mednarodnih izkušenj in prakse je posebna pozornost namenjena izgradnji spletnega arhiva. Pri tem avtorica analizira spletno gradivo in kako njegove značilnosti vplivajo na izbiro različnih metod in strategij zbiranja. Kvaliteta zajetega gradiva je odvisna od obstoječih orodij za zajemanje. Predstavljene so različne tipologije, ki temeljijo predvsem na vsebino, formo ali funkcionalnost posameznih enot gradiva, kot tudi alternativne delitve, ki jih prinaša teorija žanrov na internetu. Obravnavani so tudi problemi bibliografske obdelave in različni vidiki avtorskih pravic. Posebno pozornost avtorica daje kriterijem selekcije, ki so lahko vsebinski, formalni ali tehnični. Zelo pomemben vidik predstavljajo obstoječe strategije za trajno ohranjanje, vključno z metapodatki kot nosilci dragocenih informacij za zagotavljanje dostopa do gradiva na spletu tudi v prihodnosti.

Rezultat raziskave je celovit metodološki model, ki temelji na referenčnem modelu OAIS (Odprtega arhivskega informacijskega sistema) in ki bo postal osnova za vzpostavitev sistema za arhiviranje publikacij na spletu v NUK. Ta isti model je mogoče uporabiti tudi v drugih okoljih.

Ključne besede: spletni arhivi, izgradnja spletnih arhivov, spletne publikacije, zbiranje publikacij na medmrežju, trajno ohranjanje spletnega gradiva, metode, metodologija, modeli, OAIS.

DESARROLLO DEL ARCHIVO DE LA SLOVÉNICA EN INTERNET : UNA APROXIMACIÓN METODOLÓGICA

RESÚMEN

En 2006 el Parlamento esloveno adoptó una nueva ley de Depósito Legal que incluía las publicaciones en Internet (*www*). La Biblioteca nacional y universitaria (*Narodna in univerzitetna knjižnica - NUK*), de acuerdo a sus funciones y competencias, obtuvo a través de esta Ley la máxima responsabilidad en cuanto a la colección y archivo de los documentos electrónicos publicados en la web y que representan una parte importante de la herencia cultural eslovena.

En este sentido, los objetivos de esta tesis son los siguientes: encontrar métodos óptimos para la colección y preservación de publicaciones en la web; ajustar diferentes metodologías internacionales y crear la más apropiada para NUK; y últimamente, presentar aspectos importantes concernientes al archivo de las publicaciones electrónicas publicadas en la web que podrían contribuir en la toma de decisiones en cuanto a soluciones óptimas.

Partiendo de diferentes experiencias y prácticas internacionales, en la tesis se presta atención especial al desarrollo de archivos web. Se discuten las características de las publicaciones electrónicas publicadas en el world wide web así como su influencia en la selección de diferentes métodos y estrategias de colección. La calidad del archivo web depende de las posibilidades de las herramientas para la extracción de las publicaciones de la web. Se presentan diferentes tipologías basadas en el contenido, forma y funcionalidad de este tipo de publicaciones. Soluciones interesantes se pueden encontrar en la taxonomía introducida por la teoría de los géneros en Internet. Se discuten los problemas de descripción bibliográfica de los documentos de la web así como los aspectos de propiedad intelectual. Se presta atención especial a los criterios de selección en cuanto al contenido, forma o aspectos técnicos de las publicaciones en la web. De gran importancia son las existentes estrategias de preservación de contenidos digitales a largo plazo, incluyendo los metadatos como contenedores de valiosa información que podría asegurar el acceso a estos contenidos en el futuro.

El resultado de esta investigación es un modelo basado en el modelo referencial OAIS que fue utilizado en la implementación del sistema de archivo de NUK. Este modelo puede ser implementado también en otros entornos.

Palabras claves: archivos web, desarrollo de archivos web, publicaciones en Internet, colecciones de publicaciones en Internet, conservación a largo plazo, métodos, metodología, modelos, OAIS.

ŽIVOTOPIS

Rodila sam se u Ljubljani 26.12.1959. Nakon završetka gimnazije u Madridu (Španjolska) 1978. godine upisala sam hispanistiku, smjer španjolske književnosti, na Filozofskom fakultetu (*Facultad de filosofía y letras*) Sveučilišta Complutense u Madridu, gdje sam završila tri godine studija. Godine 1983. diplomirala sam na Filološkom fakultetu Sveučilišta u Beogradu na Odsjeku za španjolski jezik i književnost, prevoditeljski smjer, i tako stekla zvanje profesora španjolskoga jezika i književnosti.

Od 1986. do 1997. godine zaposlena sam u Međunarodnom centru za poduzeća u društvenom vlasništvu (*International Center for Public Enterprises - ICPE*) u Ljubljani. Od 1989. godine radila sam kao dokumentalistica, a poslije položenoga stručnoga ispita iz bibliotekarstva 1991. godine dodijeljeno mi je mjesto voditeljice knjižnice u Informacijsko-knjižničnoj službi ICPE-a. Iste godine upisala sam poslijediplomski studij iz informacijskih znanosti na Institutu Informacijskih Znanosti u Zagrebu. Zbog ukidanja instituta, nastavila sam studij na Fakultetu za organizaciju i informatiku Sveučilišta u Zagrebu u Varaždinu, gdje sam 1996. godine magistrirala s radom "Prestrukturiranje specijalne biblioteke u modernu informacijsku službu u Sloveniji". Magistarski je rad bio rezultat dugogodišnjega rada u izgradnji i razvoju specijalizirane knjižnice na području ekonomije, posebno menedžmenta, te aktivnoga sudjelovanja na istraživačkim projektima ICPE-a.

Od 1997. do 2001. godine zaposlena sam kao dokumentalistica u Centralnoj ekonomskoj knjižnici Ekonomskog fakulteta Sveučilišta u Ljubljani. Bila sam zadužena za elektroničke publikacije i baze podataka. Godine 2001. preuzela sam mjesto voditeljice Bibliotekarskoga istraživačkoga centra Narodne i sveučilišne knjižnice u Ljubljani (NUK). Osim koordinacije istraživanja te narodnih i međunarodnih istraživačkih projekata u NUK-u, nastavila sam sa specijalizacijom na području izgradnje digitalnih i web arhiva i dugoročne pohrane elektroničke građe. Većina mojih bibliografskih radova iz navedenih je područja.

Aktivno sudjelujem i u radu više nacionalnih i međunarodnih udruga. Od 1990. do 1996. godine bila sam individualni član FID-a, *International Federation for Information and Documentation* (FID), The Hague, Netherlands, od 1998. do 2001. godine bila sam član EBSLG-a, *European Business Schools Librarians' Group*, a od 2002. godine član je Saveza knjižničarskih društava Slovenije. Godine 2006. bila sam izabrana članica IFLA-ina odbora Sekcije za Informacijsku tehnologiju.

SADRŽAJ

	Str.
1. UVOD	1
1.1. Arhiviranje mrežne građe	1
1.2. Cilj rada	4
1.3. Metodologija	5
1.4. Organizacija rada	6
1.5. Terminologija	6
2. KARAKTERISTIKE PUBLIKACIJA NA MREŽI	9
2.1. Dubinska ili nevidljiva mreža	14
3. DIGITALNI I WEB ARHIVI	18
3.1. Definicija koncepta	18
3.2. Uvjeti za izgradnju pouzdanih digitalnih arhiva	20
3.3. Nacionalne knjižnice u ulozi pouzdanih digitalnih arhiva	22
3.4. Vrsta digitalnih repozitorija	23
3.5. Web arhivi	25
3.6. Pregled prakse i pristupa izgradnji web arhiva u svijetu	26
3.6.1. Početak izgradnje europskih web arhiva	27
3.6.2. Iskustva SAD-a u izgradnji web arhiva	31
3.6.3. Web arhiv Nacionalne knjižnice Australije	34
3.6.4. Ostala bitna iskustva u izgradnji web arhiva	39
3.6.5. Web arhivi kao rezultat akademskih istraživanja	51
3.7. Situacija u Sloveniji	53
4. ALATI I METODE PRIKUPLJANJA I POHRANE MREŽNE GRAĐE	55
4.1. Softverske aplikacije za prikupljanje mrežne građe	55

4.2. Softverske aplikacije za pristup građi	60
4.3. Razvoj u okviru konzorcija IIPC (<i>International Internet Preservation Consortium</i>)	62
4.4. Repozitorij web arhiva	72
4.5. Metode prikupljanja mrežnih stranica	73
5. TIPOLOGIJA MREŽNIH PUBLIKACIJA	80
5.1. Formalna podjela mrežne građe	83
5.2. Podjele mrežne građe na osnovi njihova sadržaja	87
6. KRITERIJI ZA ODABIR ELEKTRONIČKE GRAĐE	97
6.1. Definicija kriterija odabira mrežne građe	97
6.2. Međunarodna iskustva pri definiranju kriterija odabira	100
6.3. Tumačenje koncepta slovenike	110
6.4. Određivanje kriterija odabira u okviru slovenike	112
6.4.1. Kriteriji odabira na osnovi sadržaja	113
6.4.2. Kriteriji odabira na osnovi formalnih aspekata mrežne građe	118
6.4.3. Kriteriji odabira na osnovi tehničkih aspekata	119
6.5. Dodatno razmatranje u vezi s kriterijima odabira mrežne građe	121
7. AUTORSKO PRAVO	123
7.1. Pravo na prikupljanje mrežne građe	125
7.2. Pravo javnosti na pristup arhiviranim mrežnim stranicama	128
7.3. Pravo na dugoročnu pohranu mrežne građe	129
7.4. Zakonska rješenja arhiviranja mrežne građe u Sloveniji	129
8. BIBLIOGRAFSKA OBRADBA MREŽNE I ELEKTRONIČKE GRAĐE	132
9. DUGOROČNA POHRANA PUBLIKACIJA NA MREŽI	140
9.1. Strategije za dugoročnu pohranu elektroničke građe	146

9.1.1. Struktura elektroničke građe	146
9.1.2. Skladištenje tijeka podataka	148
9.1.3. Dugoročna pohrana pristupa elektroničkoj građi	150
9.1.3.1. Strategija čiji je cilj dugoročna pohrana strukture logičkoga zapisa	150
9.1.3.2. Strategija čiji je cilj dugoročna pohrana sadržaja zapisa, odnosno konceptualnoga objekta	153
9.2. Koncept referentnoga modela OAIS	157
9.2.1. Informacijski model	158
9.2.2. Funkcionalni model	161
9.3. Prilagođavanje repozitorija web arhiva modelu OAIS	167
10. METAPODATCI I TRAJNI IDENTIFIKATORI	171
10.1. TRAJNI IDENTIFIKATORI	188
11. IZGRADNJA METODOLOŠKOGA MODELA ZA POHRANU SLOVENSKE MREŽNE I ELEKTRONIČKE GRAĐE	193
12. ISKUSTVA NA PROJEKTIMA NUK-a	223
12.1. Ocjena rezultata projekta	236
13. ZAKLJUČAK	239
14. LITERATURA	245
POPIS TABELA	262
POPIS SLIKA	262
POPIS DODATAKA	263
UPOTRIJEBLJENE KRATICE	264
DODATCI	267
SAŽETAK, ABSTRACT, POVZETEK, RESÚMEN	289
ŽIVOTOPIS	292
SADRŽAJ	293

